

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 10, 2004, 15:29:15 ; Search time 372 Seconds  
(without alignments)  
902.171 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 456  
Sequence: 1 MCGRRPVSAAGCFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US0972032/runat\_06072004\_121451\_8379/app\_query.fasta\_1.263  
-DB=N\_Geneseq\_29Jan04 -Qfmt=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US0972032 @CGN 1.1 470 @runat\_06072004\_121451\_8379 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1980s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002s:\*  
7: Geneseq2003as:\*  
8: Geneseq2003bs:\*  
9: Geneseq2003cs:\*  
10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	990	6	ABL60606 Human ERC
C 2	164.5	36.1	597	5	AAS69040 Human encod
C 3	95	20.8	349901	9	ADC86940 Human GPC
4	94.5	20.7	24081	7	AAD54223 Streptomy
5	94.5	20.7	52101	7	AAD54217 Streptomy
C 6	94.5	20.6	696	3	AAS50885 A. sulcat
7	93.5	20.5	606	5	AAS52281 Human AFP
8	93.5	20.5	943	6	ABK70016 CDNA enco

9	93.5	20.5	943	8	ADA01379 Human PRO
10	93.5	20.5	943	8	ADA43808 Human CDN
11	93.5	20.5	943	8	ADA43576 Human CDN
12	93.5	20.5	943	8	ADA01251 Human PRO
13	93.5	20.5	943	8	ADA01135 Human CDN
14	93.5	20.5	943	8	ADA43692 Human CDN
15	93.5	20.5	943	8	ADA06954 Human PRO
16	93.5	20.5	943	8	ADA08442 Novel hum
17	93.5	20.5	943	8	ADB99735 Human PRO
18	93.5	20.5	943	8	ADB87018 Human PRO
19	93.5	20.5	943	8	ADB66173 Human PRO
20	93.5	20.5	943	9	ADB99851 Human PRO
21	93.5	20.5	943	9	ADB99506 Novel hum
22	93.5	20.5	943	9	ADB66057 Human CDN
23	93.5	20.5	943	9	ADC23455 Human CDN
24	93.5	20.5	943	9	ADC26148 Human PRO
25	93.5	20.5	943	9	ADB04975 Human PRO
26	93.5	20.5	943	9	ADB11281 Human PRO
27	93.5	20.5	943	9	ADB88212 Human PRO
28	93.5	20.5	943	9	ADD95507 Human CDN
29	93.5	20.5	943	9	ADB06437 Human PRO
30	93.5	20.5	943	9	ADB38212 Human PRO
31	93.5	20.5	943	9	ADB88328 Human PRO
32	93.5	20.5	943	9	ADD90909 Human CDN
33	93.5	20.5	943	10	ADE51762 Human CDN
34	93.5	20.5	943	10	ADE51878 Human CDN
35	93.5	20.5	943	10	ADE37736 Human CDN
36	93.5	20.5	943	10	ADE37620 Human CDN
37	93.5	20.5	943	10	ADD95391 Human CDN
38	93.5	20.5	943	10	ADE38091 Human PRO
39	93.5	20.5	943	10	ADE76180 Human PRO
40	93.5	20.5	943	10	ADE39503 Human PRO
41	93.5	20.5	943	10	ADE04307 Human PRO
42	93.5	20.5	943	10	ADE39904 Human PRO
43	93.5	20.5	943	10	ADE19769 Human PRO
44	93.5	20.5	943	10	ADE77347 Human CDN
45	93.5	20.5	943	10	ADE65455 Human PRO

ALIGNMENTS

RESULT 1  
ABL60606  
ID ABL60606 standard; cDNA; 990 BP.  
XX  
AC ABL60606;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human ERCoA3 protein encoding cDNA.  
XX  
KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;  
KW osteoporosis; cytostatic; osteopathic; human; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 203..442  
FT /\*tag= a  
FT /product= "ERCoA3 protein"

XX  
PN WO200228352-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031271.  
XX  
PR 05-OCT-2000; 2000US-0238190P.  
XX  
PA (UYCA-) UNIV CASE WESTERN RESERVE.  
XX  
PI Montano M, Sutton A;  
XX



```

XX 01-JAN-2004 (first entry)
XX Human GPCR gene SEQ ID NO:1393.
XX ds; gene; human; GPCR;
XX guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX Homo sapiens.
XX EPI270724-A2.
XX 02-JAN-2003.
XX 18-JUN-2002; 2002EP-00013517.
XX 18-JUN-2001; 2001JP-00246789.
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX Suwa M, Asai K, Akiyama Y, Aburatani H;
XX WPI; 2003-315783/31.
XX P-PSDB; ADC86941.
XX New polynucleotide, useful for preparing a composition for treating a
XX patient in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor.
XX Claim 1; SEQ ID NO 1393; 28pp; English.
XX The invention relates to a novel polynucleotide encoding a guanosine
XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX the invention may have a use in gene therapy. The polynucleotide and
XX polypeptide are useful for preparing a composition for treating a patient
XX in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor. The
XX polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
XX invention.
XX
XX Sequence 349901 BP; 106873 A; 66040 C; 67419 G; 108968 T; 0 U; 601 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3.18e+04 Length: 349901
XX Score: 95.00 Matches: 37
XX Percent Similarity: 37.39% Conservative: 6
XX Best Local Similarity: 32.17% Mismatches: 29
XX Query Match: 20.83% Indels: 43
XX DB: 9 Gaps: 5
XX
XX US-09-972-032-2 (1-79) x ADC86940 (1-349901)
XX
XX QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
XX Db 107177 GCGGGCACTGTCGGGAGACTGGGCTGC-----GCAGGAGCCACGGCAGTGGG--- 107127
XX
XX QY 23 TrpThrGlyLeuGlyGlnGlnGly-
XX Db 107126 -----GGAGGAGGCTTAGGCATGGCGGCTGTGTCTCCACGCTCCCTACAGGGAA 107073
XX
XX QY 32 -----
XX Db 107072 GCAGCTAAGCCCTGTGAGAAATCGAGCAGCAGAGCTGTGCCCGCAGTGCTAAGCCCT 107013
XX
XX QY 33 ---GlyIleGlyProGlnGlyGlnAlaSerProThrProAspCysAlaSerArgTrpPro 51
XX Db 107012 CACTGCTGGGGCAGCGGGCCCGGAGCCACGCCACACAGAACTCGCGCTGGCC 106953
XX
XX QY 52 ArgSerAlaSerArgTrpProTrpSerAla---GlyLeuThrValArgAspArgProGln 70
XX Db 106952 ACAAGCACCGCGCGAGCCCTTGGTCTCTCCACACACCTCTCCACACCTCCCGCAA 106893

```

```

QY 71 -LeuGlyCluLeuCysMet-----GlyArgGly 79
Db 106892 GCTGAGGAGCTGCTCTAGCCTCGCCAGCCAGGAGGGC 106850
XX
XX RESULT 4
XX AAD54223
XX ID AAD54223 standard; DNA; 24081 BP.
XX AC AAD54223;
XX 17-JUN-2003 (first entry)
XX Streptomyces platensis subspecies rosaceus dorrigin ORF6 DNA.
XX Polyketide biosynthesis; dorrigin; DORR; lactimidomycin; LACT; gene;
XX ds.
XX Streptomyces platensis.
XX Key Location/Qualifiers
XX CDS 1..24081
XX FT /*tag= a
XX FT /product= "ORF6 protein"
XX FT /note= "No start codon"
XX FT /partial
XX
XX WO200288176-A2.
XX 07-NOV-2002.
XX 26-APR-2002; 2002WO-CA000591.
XX 26-APR-2001; 2001US-0286346P.
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX Farnet CM, Zazopoulos E, Staffa A, Yang X;
XX WPI; 2003-201222/19.
XX P-PSDB; AAE35489.
XX
XX Novel isolated or purified polypeptide involved in biosynthesis of
XX polyketide dorrigin or polyketide lactimidomycin, useful for preparing
XX dorrigin or lactimidomycin.
XX Claim 6; Page 168-181; 312pp; English.
XX The invention relates to novel proteins involved in the biosynthesis of
XX polyketide dorrigin (DORR) or lactimidomycin (LACT) biosynthesis by
XX microorganisms. Sequences of the invention allow direct manipulation of
XX dorrigin, lactimidomycin and related chemical structures via chemical
XX engineering of the enzymes involved in the biosynthesis of dorrigin and
XX lactimidomycin. They are useful for introducing chemical handles into
XX normally inert positions that permit subsequent chemical modifications
XX and facilitate the development of polyketides. The genes and proteins of
XX the invention can also be used to generate a focused library of analogues
XX around a polyketide lead candidate to fine-tune the compound for optimal
XX properties. They are useful for generating antibodies specific for the
XX polyketide biosynthesis. The present sequence is Streptomyces platensis
XX subspecies rosaceus DORR ORF6 DNA
XX
XX Sequence 24081 BP; 3269 A; 9645 C; 8186 G; 2981 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2.41e+03 Length: 24081
XX Score: 94.50 Matches: 27
XX Percent Similarity: 42.42% Conservative: 1
XX Best Local Similarity: 40.91% Mismatches: 30
XX Query Match: 20.72% Indels: 8
XX DB: 7 Gaps: 3
XX
XX US-09-972-032-2 (1-79) x AAD54223 (1-24081)

```

QY 9 SerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-Gl 28  
 Db 12490 TCGCGCGCTGCGCTTCAGAGCGCGCACGCGCGCTCGCGAGCGCTGGACTGC 12549  
 QY 28 uGlyGlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48  
 Db 12550 CGGGTCAAGTGTCTCAACTGGGCTACTGGGCAACGTCCTCGACGAGTCTCTGGCGAC 12609  
 QY 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61  
 Db 12610 GTGACGTCCATGGGACTGGCCCGGATCGCCCGGCAACGCGATGGCGCGACTGGAGCGC 12669  
 QY 61 aGlyLeuThrValArg 66  
 Db 12670 CTCCTGGCGCGCCGC 12685

RESULT 5  
 AAD54217  
 ID AAD54217 standard; DNA; 52101 BP.  
 AC AAD54217;  
 DT 17-JUN-2003 (first entry)  
 DE Streptomyces platensis subspecies rosaceus dorrigocin DNA.  
 XX Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;  
 KW ds.  
 OS Streptomyces platensis.  
 XX Key  
 FH Location/Qualifiers  
 CDS complement(67..3720)  
 FT /\*tag= a  
 FT /product= "ORF1 protein"  
 FT 4092..5681  
 FT /\*tag= b  
 FT /product= "ORF2 protein"  
 FT /note= "No start codon"  
 FT /partial  
 FT 5767..6018  
 FT /\*tag= c  
 FT /product= "ORF3 protein"  
 FT 6023..7993  
 FT /\*tag= d  
 FT /product= "ORF4 protein"  
 FT 8009..17587  
 FT /\*tag= e  
 FT /product= "ORF5 protein"  
 FT 17634..41714  
 FT /\*tag= f  
 FT /product= "ORF6 protein"  
 FT /note= "No start codon"  
 FT /partial  
 FT 41772..47633  
 FT /\*tag= g  
 FT /product= "ORF7 protein"  
 FT /note= "No start codon"  
 FT /partial  
 FT 47635..49890  
 FT /\*tag= h  
 FT /product= "ORF8 protein"  
 FT 49922..50938  
 FT /\*tag= i  
 FT /product= "ORF9 protein"  
 FT 51234..52079  
 FT /\*tag= j  
 FT /product= "ORF10 protein"  
 XX W0200288176-A2.  
 XX 07-NOV-2002.

PF 26-APR-2002; 2002WO-CA000591.  
 XX  
 PR 26-APR-2001; 2001US-0286346P.  
 XX  
 PA (ECOP-) ECOPIA BIOSCIENCES INC.  
 XX  
 PI Farnet CM, Zazopoulos E, Staffa A, Yang X;  
 XX  
 DR WPI: 2003-20122/19  
 DR P-FSDS; AAE35488, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,  
 DR AAE35490, AAE35491, AAE35492, AAE35493.  
 XX  
 FT Novel isolated or purified polypeptide involved in biosynthesis of  
 FT polyketide dorrigocin or polyketide lactimidomycin, useful for preparing  
 FT dorrigocin or lactimidomycin.  
 XX  
 PS Claim 1; Page 85-113; 312pp; English.  
 XX  
 CC The invention relates to novel proteins involved in the biosynthesis of  
 CC polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by  
 CC microorganisms. Sequences of the invention allow direct manipulation of  
 CC dorrigocin, lactimidomycin and related chemical structures via chemical  
 CC engineering of the enzymes involved in the biosynthesis of dorrigocin and  
 CC lactimidomycin. They are useful for introducing chemical handles into  
 CC normally inert positions that permit subsequent chemical modifications  
 CC and facilitate the development of polyketides. The genes and proteins of  
 CC the invention can also be used to generate a focused library of analogues  
 CC around a polyketide lead candidate to fine-tune the compound for optimal  
 CC properties. They are useful for generating antibodies specific for the  
 CC polyketide biosynthesis. The present sequence is Streptomyces platensis  
 CC subspecies rosaceus DORR DNA  
 XX  
 SQ Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 5,21e+03 Length: 52101  
 Score: 94.50 Matches: 27  
 Percent Similarity: 42.42% Conservative: 1  
 Best Local Similarity: 40.91% Mismatches: 30  
 Query Match: 20.72% Indels: 8  
 DB: 7 Gaps: 3  
 US-09-972-032-2 (1-79) x AAD54217 (1-52101)  
 QY 9 SerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-Gl 28  
 Db 30123 TCGCGCGCTGCGCTTCAGAGCGCGCACGCGCGCTCGCGAGCGCTGGACTGC 30182  
 QY 28 uGlyGlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48  
 Db 30183 CGGGTCAAGTGTCTCAACTGGGCTACTGGGCAACGTCCTCGACGAGTCTCTGGCGAC 30242  
 QY 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61  
 Db 30243 GTGACGTCCATGGGACTGGCCCGGATCGCCCGGCAACGCGCGCTGGCGAGCGC 30302  
 QY 61 aGlyLeuThrValArg 66  
 Db 30303 CTCCTGGCGCGCCGC 30318  
 RESULT 6  
 AAA50885/C  
 ID AAA50885 standard; cDNA; 696 BP.  
 XX  
 AC AAA50885;  
 XX  
 DT 08-SEP-2000 (first entry)  
 XX  
 DE A. sulcata humanised mutant fluorescent protein, Mut1, coding sequence.  
 XX Green fluorescent protein; GFP; Mut1; protein localisation;  
 KW fluorescence resonance energy transfer; mutant; ss.  
 XX



```

OS Anemonia sulcata.
XX Key Location/Qualifiers
FH CDS 1..696
FT /*tag= a
FT /partial
FT /product= "Humanised Mut1"
XX WO200034319-A1.
XX 15-JUN-2000.
XX
XX 10-DEC-1999; 99WO-US029300.
XX
XX 11-DEC-1998; 98US-00210330.
XX 09-DEC-1999; 99US-00210330.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukyanov SA, Pradkov AF, Labas YA, Matz MV, Jiang X, Duong T;
PI Zhao X;
XX WPI; 2000-423374/36.
XX P-PSDB; AAY99734.
XX
XX Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,
PT useful for fluorescent labeling and as markers.
XX
XX Claim 8; Page 67; 71pp; English.
XX
XX The present sequence is the coding sequence for humanised mutant
CC fluorescent protein, Mut1, from non-bioluminescent Anemonia sulcata. The
CC fluorescent protein has applications in fluorescent labelling, as
CC fluorescent markers for gene expression and protein localisation studies,
CC and in fluorescence resonance energy transfer (FRET) reactions. The
CC coding sequence may be used as a source or primers and probes for
CC identifying related proteins
XX
XX Sequence 696 BP; 141 A; 257 C; 195 G; 103 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 77 Length: 696
Score: 94.00 Matches: 33
Percent Similarity: 43.02% Conservative: 4
Best Local Similarity: 38.37% Mismatches: 31
Query Match: 20.61% Indels: 18
DB: 3 Gaps: 6
US-09-972-032-2 (1-79) x AAA50885 (1-696)
QY 2 CysGlyArgProArgArgValSer---AlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
Db 592 TGGGGTGGTCCCTCGAAGCGCGGCATCTTCAGGCGGAGGGGGCGCTCTTGGAGC 533
QY 21 GlyLeuTrpThrGly-----LeuGlyGluGlyGlnGluGlyGlyProGlu 37
Db 532 GGTAGGTGGTGTGCAGGTGGCAGGTGAGTGGCGCGCGCGGCACCTTCAGGGCATCA 473
QY 38 GlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrp 57
Db 472 GGG-----ACTGGCGCGCAGCAGCAGCGCTCCACCTCGTACAGATCTCGGTGG 425
QY 58 -----ProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeu 71
Db 424 AGGGCTCCACGGCGCGCTGTCTTGTGATCA-----CGGGCGGT---CGGGCG 377
QY 72 GlyGluLeuCysMetGly 77
Db 376 GGAAGTTGTGCCAGGA 359
RESULT 7
AAH52281
ID AAH52281 standard; cDNA; 606 BP.

```

```

XX AAH52281;
XX 10-SEP-2001 (first entry)
XX Human AFP protein encoding cDNA sequence SEQ ID NO:377.
XX Human; secreted protein; secretion; bacterial cell; fungal cell;
XX eukaryotic cell; fusion protein; maltose binding protein;
XX immunoglobulin constant region; polyhistidine tag; ss.
XX Homo sapiens.
XX WO200129221-A2.
XX 26-APR-2001.
XX 20-OCT-2000; 2000WO-US029052.
XX 20-OCT-1999; 99US-0160712P.
XX (ZYMO) ZYMOGENETICS INC.
XX Conklin DC, Yee DP;
XX WPI; 2001-300340/31.
XX P-PSDB; AAG81430.
XX Isolated polypeptide for directing secretion of proteins of interest from
XX a host cell including, e.g. bacteria, includes contiguous amino acid
XX residues of polypeptide with specified amino acids.
XX Claim 9; Page 585-586; 617pp; English.
XX AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
XX to AAG81453. The secreted proteins can be used for directing the
XX secretion of proteins of interest from a host cell including bacteria,
XX fungal cells, and cultured higher eukaryotic cells. The present invention
XX also describes fusion proteins, where a secreted protein of the invention
XX is operably linked via a peptide bond or peptide linker to a second
XX protein selected from the group consisting of maltose binding protein, an
XX immunoglobulin constant region, a polyhistidine tag and a peptide given
XX in AAG81453
XX
XX Sequence 606 BP; 93 A; 198 C; 205 G; 110 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 73.3 Length: 606
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 5 Gaps: 6
US-09-972-032-2 (1-79) x AAH52281 (1-606)
QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 92 TGTGCCAGCGCTGGAGGATGGCGCGCAGCGCGCTGTGGAGGTCTCTGCTGGC 151
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyProGlu----- 37
Db 152 TGGTGGACAGGACCCCGGGAGGGCCGA-----GCCCTGGGGCGCAGAGCGCGCAGGTGG 205
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 206 ACGCACATGACTGTGAGCGCTGGGCTGGGCTCCGAGGCGCGGCTTCAGGAGTCCC 265
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 266 GAGCACCGTAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGCTGGTGGCCACGCGCG 325
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66

```



KW rheumatoid arthritis; osteopathic; antirheumatic; antiarthritic.  
 XX Homo sapiens.  
 XX  
 PN US2003068779-A1.  
 XX  
 PD 10-APR-2003.  
 XX  
 XX 16-SEP-2002; 2002US-00245107.  
 PF  
 XX 09-MAY-2001; 2001US-0290589P.  
 PR  
 PR 29-AUG-2001; 2001WO-US027099.  
 PR 18-JUL-2002; 2002US-00197942.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
 PI Fong S;  
 XX WPI; 2003-625484/59.  
 DR P-PSDB; ADA01380.  
 DR  
 XX Novel isolated PRO1313, PRO20080 or PRO21383 polypeptide useful for  
 PT stimulating proliferation of human microvascular endothelial cells, and  
 PT PRO6018 polypeptide useful for stimulating proliferation of chondrocyte  
 PT cells.  
 XX  
 XX Claim 2; Fig 111; 307pp; English.  
 PS  
 XX The invention relates to isolated human PRO polypeptides (secreted and  
 CC transmembrane polypeptides) and the polynucleotides encoding them. The  
 CC invention also relates to an antibody which specifically binds to a PRO  
 CC polypeptide, a method for stimulating the release of tumour necrosis  
 CC factor-alpha (TNF-alpha) from human blood, a method for stimulating the  
 CC proliferation or differentiation of chondrocyte cells and a method for  
 CC detecting the presence of a tumour in a mammal (e.g. adrenal, lung,  
 CC colon, breast, prostate, rectal, cervical and liver tumours). The  
 CC polynucleotides are useful in molecular biology, including uses as  
 CC hybridisation probes, in chromosome and gene mapping, in generating  
 CC antisense RNA and DNA and in gene therapy. The polynucleotides may also  
 CC be used in preparing PRO polypeptides by recombinant techniques and in  
 CC generating either transgenic animals or knock-out animals which are  
 CC useful in the development and screening of therapeutically useful  
 CC reagents. The PRO polypeptides or antibodies are used in preparing a  
 CC medicament for treating a condition responsive to the polypeptides or  
 CC antibodies, such as tumours, for stimulating and inhibiting proliferation  
 CC of human microvascular endothelial cells, for inducing endothelial cell  
 CC tube formation and for treating sports-related joint problems, articular  
 CC cartilage defects, osteoarthritis and rheumatoid arthritis. This sequence  
 CC represents a human PRO polynucleotide of the invention.  
 XX  
 SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 114 Length: 943  
 Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7  
 Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA01379 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 138 TGTGCCAGACGCTGGAGGATGGCGGCGAGCGCTGTGGAGGCTCTGTGGC 197  
 QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlyLeuGlyGlyLeuGlyGlyGlyGlyGlyGly 37  
 Db 198 TGTGTGACAGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCGGCGAGGTGG 251  
 QY 38 -----GlyGlnAlaSerProThrPro 44

Db 252 AGCACATGACTGTGAGGGCGTGGGCTCGAGGCGAGCGGCTTCAGGAGTCCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGGACCGCTCAAACTGCAGTTTCGACATGATCGCGCTGCAACCTGTGTGGCCAGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66  
 Db 372 CGCTCACCAGCGAGCGGCGCTTCCTTCCTGGGGCTGTGGGCTGCCCTGTGTGTAC 431  
 QY 66 gAspArgProGlnLeu-----GlyGluLeuCysMe 76  
 Db 432 CGAGCGCCCGTGTGTGGAGGAGGCGATGCGGTGCATTCCAACTGGCGAGTTTGTGCC 491  
 QY 76 tGly 77  
 Db 492 TGGT 495  
 RESULT 10  
 ADA43808  
 ID ADA43808 standard; cDNA; 943 BP.  
 XX  
 AC ADA43808;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human cDNA encoding secreted/transmembrane polypeptide PRO21434.  
 XX  
 KW ss; gene; human; PRO; secreted protein; transmembrane protein;  
 KW endothelial cell tube formation; chondrocyte cell differentiation;  
 KW microvascular endothelial cell; tumour; lung tumour; colon tumour;  
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;  
 KW liver tumour; cytostatic; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003064474-A1.  
 PD  
 XX 03-APR-2003.  
 PF  
 XX 16-SEP-2002; 2002US-00245859.  
 PR 29-AUG-2001; 2001WO-US027099.  
 PR 18-JUL-2002; 2002US-00197942.  
 XX  
 PA (GETH ) GENENTECH INC.  
 PI Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
 PI Fong S;  
 XX WPI; 2003-605867/57.  
 DR P-PSDB; ADA43809.  
 XX  
 PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO20080 or  
 PT PRO21383, useful in molecular biology, chromosome and gene mapping, in  
 PT generating antisense RNA and DNA, and in gene therapy.  
 XX  
 PS Claim 2; Fig 111; 308pp; English.  
 XX  
 CC The invention relates to an isolated secreted/transmembrane (PRO)  
 CC polypeptide, having at least 80% sequence identity to a sequence selected  
 CC from any one of the 57 amino acid sequences given in specification, or to  
 CC a sequence encoded by a nucleic acid molecule selected from any one of  
 CC the nucleic acids deposited under any of the ATCC accession numbers given  
 CC in specification, or a sequence having at least 80% identity to PRO  
 CC lacking its associated signal peptide, an extracellular domain of PRO  
 CC with or without its associated signal peptide. Also included are vectors,  
 CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding  
 CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by  
 CC administering PRO281, PRO1560, PRO189, PRO499, PRO6308, PRO6000,  
 CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and



US-09-972-032-2 (1-79) x ADA43576 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 138 TGTGCCAGCGCTGAGAGTGGCGGCGAGCGGCGAGCGGCTGTGAGTCTGCTGGC 197  
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyIleGlyProGlu----- 37  
 Db 198 TGTGGACAGGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGGCCAGTGG 251  
 QY 38 -----GlyGlnAlaSerProTrpPro 44  
 Db 252 ACGCAGATGACTGTAGCGCGCTGGGTGGGGCTCCGAGGAGCGCGCTTCCAGAGTCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGGCGCTGCAACCTGGTGGCCAGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGTCTACCGGAGGCGAGCTCACCTTCTCTGGGGCTGTGGGCTGCCCTGCTGTAC 431  
 QY 66 GAspArgProGlnLeu-----GlyGlnLeuCysMe 76  
 Db 432 CCGAGCGCGCTGTGGAGGAGGCCATGGCCGCTGCATTCCAACTGGGAGTTTGTCC 491

RESULT 12  
 ADA01251  
 ID ADA01251 standard; cDNA; 943 BP.  
 AC ADA01251;  
 DT 06-NOV-2003 (first entry)  
 XX Human PRO polynucleotide #56.  
 DE Human; gene; ss; PRO; secreted polypeptide; transmembrane polypeptide;  
 KW tumour necrosis factor-alpha; TNF-alpha; blood; chondrocyte cell; tumour;  
 KW cancer; lung; colon; breast; prostate; kidney; liver;  
 KW microvascular endothelial cell; endothelial cell tube formation.  
 XX Homo sapiens.  
 OS  
 PN US2003068782-A1.  
 XX  
 PD 10-APR-2003.  
 XX  
 PF 16-SEP-2002; 2002US-00245851.  
 XX  
 PR 27-APR-1999; 99US-0131271P.  
 PR 29-OCT-1999; 99US-0162506P.  
 PR 02-DEC-1999; 99WO-US028551.  
 PR 29-AUG-2001; 2001WO-US027099.  
 PR 18-JUL-2002; 2002US-00197942.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
 PI Fong S;  
 XX  
 DR WPI; 2003-625487/59.  
 DR P-FSDB; ADA01252.  
 XX  
 XX Novel isolated PRO polypeptides e.g. PRO281 and PRO1560, useful in the  
 PT preparation of a medicament for treating a condition responsive to PRO  
 PT polypeptide, and as therapeutic agents e.g. vaccines.  
 XX

Claim 2; Fig 111; 308pp; English.

PS The invention relates to isolated human PRO polypeptides (secreted and  
 XX transmembrane polypeptides) and the polynucleotides encoding them. The  
 CC invention also relates to an antibody which specifically binds to a PRO  
 CC polypeptide, a method for stimulating the release of tumour necrosis  
 CC factor-alpha (TNF-alpha) from human blood, a method for stimulating the  
 CC proliferation or differentiation of chondrocyte cells and a method for  
 CC detecting the presence of a tumour in a mammal (e.g. lung, colon, breast,  
 CC prostate, rectal, kidney and liver tumours). The polynucleotides are  
 CC useful in molecular biology, including uses as hybridisation probes, in  
 CC chromosome and gene mapping, in generating antisense RNA and DNA and in  
 CC gene therapy. The polynucleotides may also be used in preparing PRO  
 CC polypeptides by recombinant techniques and in generating either  
 CC transgenic animals or knock-out animals which are useful in the  
 CC development and screening of therapeutically useful reagents. The PRO  
 CC polypeptides or antibodies are used in preparing a medicament for  
 CC treating a condition responsive to the polypeptides or antibodies, such  
 CC as tumours, for stimulating and inhibiting proliferation of human  
 CC microvascular endothelial cells and for inducing endothelial cell tube  
 CC formation. This sequence represents a human PRO polynucleotide of the  
 CC invention.

SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 114 Length: 943  
 Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7  
 Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA01251 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 138 TGTGCCAGCGCTGAGAGTGGCGGCGAGCGGCGAGCGGCTGTGAGTCTGCTGGC 197  
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyIleGlyProGlu----- 37  
 Db 198 TGTGGACAGGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGGCCAGTGG 251  
 QY 38 -----GlyGlnAlaSerProTrpPro 44  
 Db 252 ACGCAGATGACTGTAGCGCGCTGGGGCTGGGGCTCCGAGGAGCGCGCTTCCAGAGTCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGGCGCTGCAACCTGGTGGCCAGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGTCTACCGGAGGCGAGCTCACCTTCTCTGGGGCTGTGGGCTGCCCTGCTGTAC 431  
 QY 66 GAspArgProGlnLeu-----GlyGlnLeuCysMe 76  
 Db 432 CCGAGCGCGCTGTGGAGGAGGCCATGGCCGCTGCATTCCAACTGGGAGTTTGTCC 491

RESULT 13  
 ADA01135  
 ID ADA01135 standard; cDNA; 943 BP.  
 XX  
 AC ADA01135;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human cDNA encoding secreted/transmembrane polypeptide PRO21434.  
 XX

KW ss; gene; human; PRO; secreted protein; transmembrane protein;  
 KW endothelial cell tube formation; chondrocyte cell differentiation;  
 KW microvascular endothelial cell; tumour; lung tumour; colon tumour;  
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;  
 KW liver tumour; cytostatic; vaccine.  
 XX Homo sapiens.  
 XX OS  
 XX PN US2003068780-A1.  
 XX PD 10-APR-2003.  
 XX PF 16-SEP-2002; 2002US-00245143.  
 XX 02-AUG-2000; 2000US-0222695P.  
 XX 20-JUN-2001; 2001WO-US019692.  
 XX 28-AUG-2001; 2001WO-US027099.  
 XX 18-JUL-2002; 2002US-00197942.  
 XX (GETH ) GENENTECH INC.  
 XX PA Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
 PI Pong S;  
 XX WPI; 2003-625485/59.  
 DR P-PSDB; ADA01136.  
 XX Novel isolated PRO polypeptides e.g. PRO281 and PRO1560, useful in the  
 PT preparation of a medicament for treating a condition responsive to PRO  
 PT polypeptide, and as therapeutic agents e.g. vaccines.  
 XX Claim 2; Fig 111; 307pp; English.  
 PS  
 XX The invention relates to an isolated secreted/transmembrane (PRO)  
 CC polypeptide, having at least 80% sequence identity to a sequence selected  
 CC from any one of the 57 amino acid sequences given in specification, or to  
 CC a sequence encoded by a nucleic acid molecule selected from any one of  
 CC the nucleic acids deposited under any of the ATCC accession numbers given  
 CC in specification, or a sequence having at least 80% identity to PRO  
 CC lacking its associated signal peptide, an extracellular domain of PRO  
 CC with or without its associated signal peptide. Also included are vectors,  
 CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding  
 CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by  
 CC administering PRO281, PRO1560, PRO189, PRO4499, PRO6308, PRO6000,  
 CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and  
 CC an oligonucleotide probe derived from any one of the above nucleotide  
 CC sequences. PRO6018 polypeptide is useful for stimulating the  
 CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080  
 CC and PRO21383 polypeptides are useful for stimulating the proliferation of  
 CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006  
 CC polypeptides are useful for inhibiting the proliferation of human  
 CC microvascular endothelial cells. PRO polypeptides are useful for  
 CC detecting the presence of tumour in a mammal, including tumours of lung,  
 CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,  
 CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO20933 and  
 CC PRO34274 polypeptides are useful for inducing endothelial cell tube  
 CC formation. PRO or the antibody are useful in the preparation of a  
 CC medicament for treating a condition responsive to PRO polypeptide. The  
 CC oligonucleotide probes are useful for isolating genomic and cDNA  
 CC nucleotide sequences, for measuring or detecting the expression of an  
 CC associated gene, and as antisense probes. PRO nucleic acid is useful as a  
 CC hybridisation probe, in chromosome and gene mapping, in the generation of  
 CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and  
 CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The  
 CC present sequence encodes a PRO protein.  
 XX SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 114 Length: 943  
 Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7

Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 8 Gaps: 6  
 US-09-972-032-2 (1-79) x ADA01135 (1-943)  
 QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGGCGAGCGCTGTGGAGGTCTCTGTGGC 137  
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyViledGlyProGlu----- 37  
 Db 198 TGTGGACAGGACCCGGGAGGGCCGA-----GCCCTGGGGCCAGAGCGGCGAGGTGG 251  
 QY 38 -----GlyGlnAlaSerProThrPro 44  
 Db 252 ACGCACATGACTGTGAGCGCTGGGCTGGGCTCCGAGGCGAGCGGCTTCCAGGAGTCCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGCACCGTCAAACTGACGTTGACATGATGCGGCGCTCAACCTGGTGGCCACGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGCTCACCGCAGCGCAGCTCACCTTCTCTCTGGGGCTGGTGGGCTGCCCCCTGCTGTAC 431  
 QY 66 gAspArgProGlnLeu-----GlyGluLeuCysMe 76  
 Db 432 CCGACGCGCGCGCTGTGGGAGGAGGCCATGGCCGCTGCATTCACACTGGCGAGTTTGTTC 491  
 QY 76 tGly 77  
 Db 492 TGGT 495  
 RESULT 14  
 ADA43692  
 ID ADA43692 standard; cDNA; 943 BP.  
 XX AC ADA43692;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Human cDNA encoding secreted/transmembrane polypeptide PRO21434.  
 KW ss; Gene; human; PRO; secreted protein; transmembrane protein;  
 KW endothelial cell tube formation; chondrocyte cell differentiation;  
 KW microvascular endothelial cell; tumour; lung tumour; colon tumour;  
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;  
 KW liver tumour; cytostatic; vaccine.  
 XX OS Homo sapiens.  
 XX PN US2003073190-A1.  
 XX PD 17-APR-2003.  
 XX 09-SEP-2002; 2002US-00238283.  
 XX 01-JUL-1998; 98US-0091358P.  
 XX 02-JUN-1999; 99WO-US012252.  
 XX 20-JUL-1999; 99US-0144758P.  
 XX 28-JUL-1999; 99US-0146222P.  
 XX 25-AUG-1999; 99US-00380137.  
 XX 30-MAR-2000; 2000WO-US008439.  
 XX 02-JUN-2000; 2000WO-US015264.  
 XX 29-AUG-2001; 2001WO-US027099.  
 XX 18-JUL-2002; 2002US-00197942.  
 XX (GETH ) GENENTECH INC.  
 XX PA Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
 PI Pong S;

XX WPI; 2003-585304/55.  
DR P-PSDB; ADA43693.  
XX  
PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO20080 or  
PT PRO21383, useful in molecular biology, chromosome and gene mapping, in  
PT generating antisense RNA and DNA, and in gene therapy.  
XX  
XX  
PS Claim 2; Fig 111; 352pp; English.  
XX  
CC The invention relates to an isolated secreted/transmembrane (PRO)  
CC polypeptide, having at least 80% sequence identity to a sequence selected  
CC from any one of the 57 amino acid sequences given in specification, or to  
CC a sequence encoded by a nucleic acid molecule selected from any one of  
CC the nucleic acids deposited under any of the ATCC accession numbers given  
CC in specification, or a sequence having at least 80% identity to PRO  
CC lacking its associated signal peptide, an extracellular domain of PRO  
CC with or without its associated signal peptide. Also included are vectors,  
CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding  
CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by  
CC administering PRO281, PRO1560, PRO189, PRO4499, PRO6308, PRO6000;  
CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and  
CC an oligonucleotide probe derived from any one of the above nucleotide  
CC sequences. PRO6018 polypeptide is useful for stimulating the  
CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080  
CC and PRO21383 polypeptides are useful for stimulating the proliferation of  
CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006  
CC polypeptides are useful for inhibiting the proliferation of human  
CC microvascular endothelial cells. PRO polypeptides are useful for  
CC detecting the presence of tumour in a mammal, including tumours of lung,  
CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,  
CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO21207, PRO20933 and  
CC PRO34274 polypeptides are useful for inducing endothelial cell tube  
CC formation. PRO or the antibody are useful in the preparation of a  
CC medicament for treating a condition responsive to PRO polypeptide. The  
CC oligonucleotide probes are useful for isolating genomic and cDNA  
CC associated sequences, for measuring or detecting the expression of an  
CC associated gene, and as antisense probes. PRO nucleic acid is useful as a  
CC hybridisation probe, in chromosome and gene mapping, in the generation of  
CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and  
CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The  
CC present sequence encodes a PRO protein.  
XX  
SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 114 Length: 943  
Score: 93.50 Matches: 36  
Percent Similarity: 35.25% Conservative: 7  
Best Local Similarity: 29.51% Mismatches: 31  
Query Match: 20.50% Indels: 48  
DB: Gaps: 6

US-09-972-032-2 (1-79) x ADA43692 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
DB 138 TGTGCCAGACGCTGGAGGATGGCGAGCGCGAGCTGGGGCTGTGGAGTCTGTGGTGC 197  
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyProGlu----- 37  
DB 198 TGTGTGACAGACCCGGGGAGGGCCGA-----GCCCTGGGGCCAGAGCGCGGCGAGTCC 251  
QY 38 -----GlyGlnAlaSerProThrPro 44  
DB 252 ACGCACATGACTGTAGGCGCTGGGCTGGGGCTCCGAGGAGCGCGGCTTCAGGAGTCCC 311  
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
DB 312 GAGGACCGTCAAACTGCAGTTCGACATGATGGCGCCCTGCAACCTGGGGCCAGCGCG 371  
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66

Db 372 CGCTCACCGCAGGCGAGCTCACCTTCCTCTGGGGCTGGTGGGGCTGCCCTGTGTGTAC 431  
QY 66 gAspArgProGlnLeu-----GlyGluLeuCysMe 76  
DB 432 CGGAGCGCCCGTCTGTGGGAGGAGGCGCATGCGCGTGCATTCCAACTGCGGAGTTTGTCC 491  
QY 76 tGly 77  
DB 492 TGGT 495

RESULT 15  
ADA06954  
ID ADA06954 standard; cDNA; 943 BP.  
XX  
AC ADA06954;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human PRO polynucleotide #56.  
XX  
KW Human; PRO; gene; ss; secreted polypeptide; transmembrane polypeptide;  
KW tumour necrosis factor-alpha; TNF-alpha; blood; chondrocyte cell; tumour;  
KW adrenal; lung; colon; breast; prostate; rectum; cervix; liver;  
KW microvascular endothelial cell; endothelial cell tube formation;  
KW bone disorder; cartilage disorder; sports injury; proteoglycan;  
KW cartilage; sports-related joint problem; articular cartilage defect;  
KW osteoarthritis; rheumatoid arthritis; haemoglobin-associated disorder;  
KW thalassemia; immune system cell infiltration; cancer; vulnery;  
KW antianemic; osteopathic; antirheumatic; antiarthritic.  
XX  
OS Homo sapiens.  
XX  
PN US2003068781-A1.  
XX  
PD 10-APR-2003.  
XX  
PF 16-SEP-2002; 2002US-00245771.  
XX  
PR 03-AUG-1999; 99US-0146843P.  
PR 15-MAY-2000; 2000WO-US013358.  
PR 29-AUG-2001; 2001WO-US027099.  
PR 18-JUL-2002; 2002US-00197942.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;  
PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;  
PI Fong S;  
XX  
DR WPI; 2003-625486/59.  
DR P-PSDB; ADA06955.  
XX  
PT Novel secreted and transmembrane polypeptides, PRO polypeptides useful  
PT for stimulating proliferation or differentiation of chondrocyte cells and  
PT inducing endothelial cell tube formation.  
XX  
PS Claim 2; Fig 111; 307pp; English.  
XX  
CC The invention relates to isolated human PRO polypeptides (secreted and  
CC transmembrane polypeptides) and the polynucleotides encoding them. The  
CC invention also relates to an antibody which specifically binds to a PRO  
CC polypeptide, a method for stimulating the release of tumour necrosis  
CC factor-alpha (TNF-alpha) from human blood, a method for stimulating the  
CC proliferation or differentiation of chondrocyte cells and a method for  
CC detecting the presence of a tumour in a mammal (e.g. adrenal, lung,  
CC colon, breast, prostate, rectal, cervical and liver tumours). The  
CC polynucleotides are useful in molecular biology, including uses as  
CC hybridisation probes, in chromosome and gene mapping, in generating  
CC antisense RNA and DNA and in gene therapy. The polynucleotides may also  
CC be used in preparing PRO polypeptides by recombinant techniques and in  
CC generating either transgenic animals or knock-out animals which are  
CC useful in the development and screening of therapeutically useful  
CC reagents. The PRO polypeptides or antibodies are used in preparing a

CC medicament for treating a condition responsive to the polypeptides or  
 CC antibodies, such as tumours, for stimulating and inhibiting proliferation  
 CC of human microvascular endothelial cells, for inducing endothelial cell  
 CC tube formation and for treating various bone and/or cartilage disorders  
 CC such as sports injuries and arthritis. PRO polypeptides which stimulate  
 CC the release of proteoglycans from cartilage are useful for treating  
 CC osteoarthritis and rheumatoid arthritis. PRO polypeptides are also useful  
 CC for treating various mammalian haemoglobin-associated disorders such as  
 CC various thalassaemias and conditions which may benefit from enhanced  
 CC local immune system cell infiltration. This sequence represents a human  
 CC PRO polynucleotide of the invention.

XX  
 SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 114 Length: 943  
 Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7  
 Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA06954 (1-943)

Qy 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCGTGGGGCTGTGGAGGTCTCTGCTGC 197  
 Qy 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 37  
 Db 198 TGGTGGACAGACCCGGGGAGGGCCGA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251  
 Qy 38 -----GlyGlnAlaSerProThrPro 44  
 Db 252 ACGCACATGACTGTGAGGCGCTGGGCTGGGGCTCCGAGGAGCGCGCTTCCAGGAGTCCC 311  
 Qy 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGGACCCGTCAAACTGCAGTTCGATGATGCGCGCCTGCAACTGTGGTGGCCACGCGCG 371  
 Qy 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGCTCACCGCAGCGCAGCTACCTTCTCTGGGGCTGGTGGGCTGCCCCCTGCTGTCTAC 431  
 Qy 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76  
 Db 432 CCGACGCCCGCTGTGGAGAGAGGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTGCC 491  
 Qy 76 tGly 77  
 Db 492 TGGT 495

Search completed: July 10, 2004, 16:58:44  
 Job time : 452 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 13, 2004, 08:38:15 ; Search time 2883 Seconds  
(without alignments)  
818.284 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 79  
Sequence: 1 MCGPRVSGCGFADAHWT.....SAGLTVRDRLGELCMGRG 79

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55023952

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09972032/runat\_06072004.121534.8884/app query.fasta\_1.263  
-DB=EST -QPMI=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=plo  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000.  
-USPR=US09972032 @CNG 1.1 3437 @runat\_06072004.121534.8884 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pin.\*  
20: em\_gss\_vit.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mas.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_pbg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	79	100.0	1467	11	BC039500 Homo sapi
2	45	57.0	691	13	BU620427 UI-H-FL1-
3	44	55.7	1201	13	EX406138 BX406138
4	34	43.0	669	12	BM679577 UI-E-E00-
5	34	43.0	694	13	BU684594 UI-CF-EN1
6	34	43.0	760	12	BM980506 UI-CF-EN1
7	33	41.8	632	13	BU617331 UI-H-DF0-
8	33	41.8	633	14	CA313142 UI-CF-FN0
9	32	40.5	623	14	CD742581 UI-H-FT2-
10	23	29.1	628	13	EQ447041 UI-H-EU1-
11	22	27.8	1084	12	BM806108 AGENCOURT
12	17	21.5	282	10	BF364571 PMI-NM108
13	9	11.4	163	14	R86559 RABEST126T
14	9	11.4	181	9	AA656691 VP9SC01.R
15	9	11.4	424	9	AA593543 m28C03.s
16	9	11.4	526	9	AI022772 cw99b10.x
17	9	11.4	554	14	CB471798 sn45_A09.
18	9	11.4	627	13	BU405769 604138917
19	9	11.4	691	13	BU113858 603131524
20	9	11.4	750	13	BU403370 604138868
21	9	11.4	793	13	BU357204 603475234
22	9	11.4	838	13	BU354677 603475426
23	9	11.4	1047	10	BE621897 601494186
24	9	11.4	1076	10	BF168132 601776390
25	8	10.1	134	28	AZ099785 RPT2-23-4
26	8	10.1	172	10	AW384572 IL2-HT039
27	8	10.1	172	14	W30123 mc27B09.r1
28	8	10.1	176	13	EQ761225 EBRO04.SQ
29	8	10.1	187	28	BH23431 1006113A0
30	8	10.1	197	14	W14247 mb27f07.r1
31	8	10.1	201	29	CE017568 tigr-Gss-
32	8	10.1	215	10	BB293476 BB293476
33	8	10.1	218	29	CE660402 tigr-gss-
34	8	10.1	220	9	AI005741 ua74h11.r
35	8	10.1	248	28	AQ113700 CIT-HSP-2
36	8	10.1	252	28	AZ577620 12a02 Sho
37	8	10.1	272	10	BF721038 mab62a09.
38	8	10.1	272	13	BQ807616 NISC_Kk06
39	8	10.1	279	10	BF651626 274459 MA
40	8	10.1	279	10	BF012245 u640a01.Y
41	8	10.1	280	14	F27454 HSPDI5401.H
42	8	10.1	283	10	BE717721 RC4-HT078
43	8	10.1	283	10	BE837703 RC2-FN009
44	8	10.1	284	14	CD810225 957 Lingu
45	8	10.1	286	10	BE837692 RC2-FN009

# ALIGNMENTS

RESULT 1  
BC039500  
LOCUS BC039500 1467 bp mRNA linear HTC 04-MAR-2003  
DEFINITION Homo sapiens, clone IMAGE:5555626, mRNA.  
ACCESSION BC039500  
VERSION BC039500.1 GI:25058499  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1467)  
AUTHORS Strausberg,R.



```

VERSION      BX406138.1  GI:30648317
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 1201)
AUTHORS      Mammalia; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE        Mammalia; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL      Normalization and subtraction: two approaches to facilitate gene
COMMENT      Normalization and subtraction: two approaches to facilitate gene
              discovery
              Genome Res. 6 (9), 791-806 (1996)
              97044477
              889548
              Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com).
              Seq primer: M13 Forward
              POLYA=Yes.
              Location/Qualifiers
                1..669
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-E-E00-aia-1-05-0-UI"
                /tissue_type="fetal eye"
                /dev_stage="fetal"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /clone_lib="UI-E-E00"
                /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoR I; Site_2: Not I;
                UI-E-E00 is a cDNA library containing the following
                tissue(s): fetal eye. The library was constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into pT7T3-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (dT)18 tail. The
                sequence tag for this library is CCGGTATACC. This library
                was created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI).
                TAG_TISSUE=human fetal eye
                TAG_LIB=UI-E-E00
                TAG_SEQ=CCGGTATACC"

FEATURES             source
    source
    Location/Qualifiers
      1..669
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CS0DM010YD19"
      /tissue_type="FETAL LIVER"
      /dev_stage="fetal"
      /clone_lib="Homo sapiens FETAL LIVER"
      /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
      was primed with a NotI-oligo(dT) primer. Five prime end
      enriched, double-strand cDNA was digested with Not I and
      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
      vector. Library was not normalized."

ORIGIN
Alignment Scores:
Pred. No.:      5,05e-29      Length:      1201
Score:          44.00         Matches:      65
Percent Similarity: 98.48%      Conservative: 0
Best Local Similarity: 98.48%      Mismatches: 1
Query Match:     55.70%         Indels:      1
DB:              13            Gaps:        0

US-09-972-032-2 (1-79) x BX406138 (1-1201)

Qy 14 PhenAlaSerArgTrpThrGlyLeuThrThrGlyLeuGlyGlnGlyGly 33
Db 674 TTCGCTGATGCACATTGGACCGGCTCTGGACTGGGCTAGGGAAGGCGAGGCGGA 615
Qy 34 IleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSer 53
Db 614 ATTGG-CCCGAGGGCCAGGCTCCGCCACCCCGACTGGCCCTCCCGTGGCCCGCAGC 556
Qy 54 AlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGlu 73
Db 555 GCCCTCCCGGTGGCCCTGGAGTGCAGGCTTTACCGTCCGAGATGTCGCGCAACTGGGCGAG 496
Qy 74 LeuCysMetGlyArgGly 79
Db 495 CTGTGTCATGGGGCGTGGC 478

RESULT 4
BM679577/c
LOCUS          BM679577
DEFINITION    UI-E-E00-aia-1-05-0-UI.s1 UI-E-E00 Homo sapiens cDNA clone
              669 bp mRNA linear EST 27-FEB-2002
ACCESSION     BM679577
VERSION       BM679577.1 GI:18989473
KEYWORDS      EST.
SOURCE        Homo sapiens (human)

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eukarya; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 669)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
              Genome Res. 6 (9), 791-806 (1996)
              97044477
              889548
              Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com).
              Seq primer: M13 Forward
              POLYA=Yes.
              Location/Qualifiers
                1..669
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-E-E00-aia-1-05-0-UI"
                /tissue_type="fetal eye"
                /dev_stage="fetal"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /clone_lib="UI-E-E00"
                /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoR I; Site_2: Not I;
                UI-E-E00 is a cDNA library containing the following
                tissue(s): fetal eye. The library was constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into pT7T3-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (dT)18 tail. The
                sequence tag for this library is CCGGTATACC. This library
                was created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI).
                TAG_TISSUE=human fetal eye
                TAG_LIB=UI-E-E00
                TAG_SEQ=CCGGTATACC"

ORIGIN
Alignment Scores:
Pred. No.:      2,33e-20      Length:      569
Score:          34.00         Matches:      34
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     43.04%         Indels:      0
DB:              12            Gaps:        0

US-09-972-032-2 (1-79) x BM679577 (1-669)

Qy 46 CysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrVal 65
Db 642 TCGCGCTCCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGGCTTACCGTCC 583
Qy 66 ArgAspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79
Db 582 CGAGATCGTCCGCAACTGGCGAGCTGTGTCATGGGGCGGTGGC 541

```

```

RESULT 5
BU684594/c
LOCUS
DEFINITION
  UI-CF-EN1-act-a-22-0-UI.s1 694 bp mRNA linear EST 07-OCT-2002
  UI-CF-EN1-act-a-22-0-UI 3', mRNA sequence.
ACCESSION
  BU684594
VERSION
  BU684594.1 GI:23537704
SOURCE
  EST.
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
  97044477
  8889548
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.

FEATURES
  source
  1..694
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="UI-CF-EN1-act-a-22-0-UI"
  /tissue_type="Primary Lung Cystic Fibrosis Epithelial
  Cells"
  /dev_stage="Adult"
  /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
  /clone_lib="UI-CF-EN1"
  /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
  modified polylinker; Site 1: EcoR I; Site 2: Not I;
  UI-CF-EN1 is a normalized cDNA library containing the
  following tissue(s): Primary Lung Cystic Fibrosis
  Epithelial Cells. The library was constructed according to
  Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
  1996. First strand cDNA synthesis was primed with an
  oligo-dT primer containing a Not I site. Double stranded
  cDNA was ligated to an EcoR I adaptor, digested with Not
  I, and cloned directionally into pT7T3-Pac vector. The
  oligonucleotide used to prime the synthesis of
  first-strand cDNA contains a library tag sequence that is
  located between the Not I site and the (dT)18 tail. The
  sequence tag for this library is CTCCTCAGGT.
  TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
  6hr to LPS 24h
  TAG LIB=UI-CF-EN1
  TAG_SEQ=CTGCTCAGGT"

ORIGIN
Alignment Scores:
Pred. No.: 2,42e-20 Length: 694
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.04% Indels: 0
DB: 13 Gaps: 0

```

```

US-09-972-032-2 (1-79) x BU684594 (1-694)
QY 46 CysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrVal 65
DB 647 TCGCGCTCCGGTGGCCCGCAGCGCCCTCCGGTGGCCCTCCGGTGGCAGGTCTTACCGTC 588
QY 66 ArgAspArgProGlnLeuGlyGlyLeuGlyMetGlyArgGly 79
DB 587 CGAGATGCTCCGCACTGGCGAGCTGTGTCATGGGCGCGTGC 546

RESULT 6
LOCUS
DEFINITION
  UI-CF-EN1-add-a-08-0-UI.s1 760 bp mRNA linear EST 21-FEB-2003
  UI-CF-EN1-add-a-08-0-UI 3', mRNA sequence.
ACCESSION
  BM980506
VERSION
  BM980506.1 GI:19602038
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
  97044477
  8889548
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.

FEATURES
  source
  1..760
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="UI-CF-EN1-add-a-08-0-UI"
  /tissue_type="Primary Lung Cystic Fibrosis Epithelial
  Cells"
  /dev_stage="Adult"
  /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
  /clone_lib="UI-CF-EN1"
  /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
  modified polylinker; Site 1: EcoR I; Site 2: Not I;
  UI-CF-EN1 is a normalized cDNA library containing the
  following tissue(s): Primary Lung Cystic Fibrosis
  Epithelial Cells. The library was constructed according to
  Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
  1996. First strand cDNA synthesis was primed with an
  oligo-dT primer containing a Not I site. Double stranded
  cDNA was ligated to an EcoR I adaptor, digested with Not
  I, and cloned directionally into pT7T3-Pac vector. The
  oligonucleotide used to prime the synthesis of
  first-strand cDNA contains a library tag sequence that is
  located between the Not I site and the (dT)18 tail. The
  sequence tag for this library is CTCCTCAGGT.
  TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
  6hr to LPS 24h

```

TAG\_LIB=UI-CF-ENI  
TAG\_SEQ=CTGCTCAGGT

## ORIGIN

Alignment Scores: 2.66e-20 Length: 760  
Pred. No.: 34.00 Matches: 65  
Score: 97.01% Conservative: 0  
Percent Similarity: 97.01% Mismatches: 1  
Best Local Similarity: 97.01% Indels: 2  
Query Match: 43.04% Gaps: 0  
DB: 12

US-09-972-032-2 (1-79) x BM980506 (1-760)

Qy 14 PheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlnGlyGly 33  
Db 732 TTCGCTGATGACATTTGGACCGGCTCTGACTGGGCTAGGGAGGAGGAGGCGGA 673  
Qy 34 IleGlyProGlnGlyGlnAlaSerProThr-ProAspCysAlaSerArgTrpProArgSe 53  
Db 672 ATTGGGCGGAGGCGCAGGCTCGCCGACCCCGAN-TGGCCCTCCGGTGGCCCGCAG 614  
Qy 53 rAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGly 73  
Db 613 CGCTCCCGGTGGCCCTGGAGTGCAGTCTTACCGTCGAGATCGTCCGCAACTGGCGCA 554  
Qy 73 uLeuCysMetGlyArgGly 79  
Db 553 GCTGTGCATGGGCGGTGGC 535

## RESULT 7

BU617331/c  
LOCUS BU617331 632 bp mRNA linear EST 23-SEP-2002  
DEFINITION UI-H-DF0-bep-n-09-0-UI.s1 NCI\_CGAP\_DFO Homo sapiens cDNA clone  
UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.  
ACCESSION BU617331  
VERSION BU617331.1 GI:23283539  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 632)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

Location/Qualifiers  
1..632  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-DF0-bep-n-09-0-UI"  
/tissue\_type="Subchondral Bone"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI\_CGAP\_DFO"  
/note="Organ: Bone; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site: 1; Ecor I; Site: 2; Not 1;  
NCI CGAP DFO is a cDNA library containing the following  
tissue(s): Subchondral Bone. The library was constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pT73-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is GTTAAGCGTC.

TAG\_TISSUE=Subchondral bone  
TAG\_LIB=UI-H-DF0  
TAG\_SEQ=GTTAAGCGTC

## ORIGIN

Alignment Scores: 1.71e-19 Length: 632  
Pred. No.: 33.00 Matches: 33  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 41.77% Gaps: 0  
DB: 13

US-09-972-032-2 (1-79) x BU617331 (1-632)

Qy 47 AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66  
Db 630 GCCTCCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGAGTGCAGGTCTTACCGTCCGA 571  
Qy 67 AspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79  
Db 570 GATCGTCGCAACTGGCGAGCTGTGCAITGGGCGTGGC 532

## RESULT 8

CA313142/c  
LOCUS CA313142 633 bp mRNA linear EST 04-NOV-2002  
DEFINITION UI-CF-FNO-aex-n-23-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone  
UI-CF-FNO-aex-n-23-0-UI 3', mRNA sequence.  
ACCESSION CA313142  
VERSION CA313142.1 GI:24531240  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 633)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: [paul-mccray@uiowa.edu](mailto:paul-mccray@uiowa.edu)  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems  
([www.openbiosystems.com](http://www.openbiosystems.com)).  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

Location/Qualifiers  
1..633  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-FNO-aex-n-23-0-UI"  
/tissue\_type="Human Lung Epithelial cells"  
/lab\_host="DH10B (Life Technologies)" (T1 phage resistant)"

/clone\_lib="UI-CF-FNO"  
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: Ecor I; Site 2: Not I;  
 UI-CF-FNO is a subtracted cDNA library derived from two  
 normalized Human lung epithelial cell libraries (EN1 and  
 DUL). The library was subtracted according to Bernaldo,  
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. For additional information, contact:  
 bento-soares@uiowa.edu  
 TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
 6hr to LPS 24h  
 TAG LIB=UI-CF-FNO  
 TAG\_SEQ=CTGCTCAGGT"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.72e-19 Length: 633  
 Score: 33.00 Matches: 33  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 41.77% Indels: 0  
 DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CA313142 (1-633)

Qy 47 AlasArGrpProArGrSerAlasArGrTrpProTrpSerAlaglyLeuThrValArg 66  
 |||||  
 Db 633 GCCTCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCAGGTCTTACCGTCCGA 574  
 |||||  
 Qy 67 AspArGrProGlnLeuGlyGluLeuCysMetGlyArgGly 79  
 |||||  
 Db 573 GATCGTCCGACTGGCGAGCTGTGATGGCGGTGGC 535  
 |||||

## RESULT 9

CD742581/c  
 LOCUS  
 DEFINITION UI-H-FT2-bj1-i-10-0-UI.s1 NCI CGAP FT2 Homo sapiens cDNA clone  
 UI-H-FT2-bj1-i-10-0-UI 3', mRNA sequence.  
 CD742581  
 VERSION  
 KEYWORDS EST.  
 SOURCE CD742581.1 GI:32293431  
 Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 623)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Dr. Gary W. Hunninghake, U of I  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/cgap.html  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

source  
 1..623  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FT2-bj1-i-10-0-UI"  
 /tissue\_type="Aveolar Macrophage"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP FT2"  
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: Ecor I; Site 2: Not I;  
 NCI CGAP FT2 is a subtracted cDNA library constructed from

a pool of 81 RNA samples from Alveolar Macrophages  
 challenged with different treatments. The library was  
 subtracted according to Bernaldo, Lennon and Soares, Genome  
 Research, 6:791-806, 1996. The tissue was provided by Dr.  
 Gary W. Hunninghake of the University of Iowa.  
 TAG TISSUE=Human Lung Aveolar Macrophage  
 TAG LIB=UI-H-FT2  
 TAG\_SEQ=GGCATGCGC"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.32e-18 Length: 623  
 Score: 32.00 Matches: 32  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 40.51% Indels: 0  
 DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CD742581 (1-623)

Qy 48 SerArGrpProArGrSerAlasArGrTrpProTrpSerAlaglyLeuThrValArg 67  
 |||||  
 Db 622 TCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCAGGTCTTACCGTCCGAGAT 563  
 |||||  
 Qy 68 ArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79  
 |||||  
 Db 562 CGTCCGAACCTGGCGAGCTGTGATGGCGGTGGC 527  
 |||||

## RESULT 10

BQ447041/c  
 LOCUS  
 DEFINITION UI-H-EUI-bac-p-06-0-UI.s1 NCI CGAP Ctl Homo sapiens cDNA clone  
 UI-H-EUI-bac-p-06-0-UI 3', mRNA sequence.  
 BQ447041  
 ACCESSION  
 VERSION  
 KEYWORDS EST.  
 SOURCE BQ447041.1 GI:21250153  
 Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 628)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

source  
 1..628  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-EUI-bac-p-06-0-UI"  
 /tissue\_type="Osteoarthritic Cartilage"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP Ctl"  
 /note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: Ecor I; Site 2: Not I;  
 NCI CGAP Ctl is a normalized cDNA library containing the  
 following tissue(s): Osteoarthritic Cartilage The library  
 was constructed according to Bernaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an Ecor I

adaptor, digested with Not I, and cloned directionally into p77N3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is TGATCACGCT.

TAG TISSUE=osteoarthritic cartilage  
TAG LIB=UI-H-EUI  
TAG\_SEQ=TGATCACGCT"

ORIGIN

Alignment Scores:

Pred. No.: 1.46e-10 Length: 629  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 29.11% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BQ447041 (1-628)

QY 57 TtrProtrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMet 76  
DB 601 TGGCCCTGGAGTGCAGCTCTTACCGTCGAGATCGTCGCACTGGCGGAGCTGTGCATG 542

QY 77 GlyArgGly 79  
DB 541 GGGGCTGGC 533

RESULT 11

BM806108

LOCUS BM806108 1084 bp mRNA linear EST 05-MAR-2002  
DEFINITION AGENCOURT\_6553891 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5555626  
5', mRNA sequence.

ACCESSION BM806108.1 GI:19122931

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1084)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHAM12276 row: b column: 11

High quality sequence stop: 672.

Location/Qualifiers

FEATURES

source

1..1084

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5555626"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb. "

ORIGIN

Alignment Scores:

Pred. No.: 2.05e-09 Length: 1084  
Score: 22.00 Matches: 22

Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 27.95% Indels: 0  
DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM806108 (1-1084)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20  
DB 672 ATGTGTGGAGACCGCTCGGTAAGCGCTTGGTGGCTTCGCTGATGCATGGACC 731

QY 21 GlyLeu 22  
DB 732 GGGCTC 737

RESULT 12

LOCUS

BF364571/c

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

SEQUENCE TAGS

PROC. NATL. ACAD. SCI. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM1&t2=PM1-NN1084-

240900-005-G05&t3=2000-09-24&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 27

High quality sequence stop: 282.

Location/Qualifiers

1..282

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="NN1084"

/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

ORIGIN

Alignment Scores:

Pred. No.: 1.43e-05 Length: 282  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservativeness: 0





ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 424)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-[email.nih.gov](mailto:email.nih.gov)  
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Stratagene, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 418.  
 FEATURES  
 Location/Qualifiers  
 1..424  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1085188"  
 /tissue\_type="gastric tumor"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="NCI CGAP Gas1"  
 /note="Organ: stomach; Vector: Bluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Pooled gastric tumors. 5' adaptor sequence: 5'  
 GAATTCGGCACGAG 3' 3' adaptor sequence: 5'  
 CTCGAGTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

# ORIGIN

Alignment Scores:  
 Pred. No.: 310 Length: 424  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 9 Gaps: 0

US-09-972-032-2 (1-79) x AA593543 (1-424)

QY 23 TrpThrGlyLeuGlyGluGlyGlnGlu 31  
 |||||  
 Db 374 TGGACAGGGCTCGAGAGGCCAGGAG 400

Search completed: July 13, 2004, 15:53:46  
 Job time : 2893 secs



```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      July 12, 2004, 18:17:11 ; Search time 403 Seconds
           (without alignments)
           955.426 Million cell updates/sec

Title:
Perfect score: 456
Sequence:      1 MCGRRVRSAGCGFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop  6.0 , Delext  7.0

Searched:      3183909 seqs, 2436941669 residues

Total number of hits satisfying chosen parameters:      6367818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xih
-Q=/cgn2_1/USPTO.spool/US0972032/runat_06072004_121455_8495/app_query.fasta_1.263
-DB=published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0972032@cgn_1_1_511@runat_06072004_121455_8495
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
------------	-------------	--------------	-------	-------------

	1	456	100.0	990	13	US-09-972-032-1	Sequence 1, Appli
	2	95	20.8	3642	16	US-10-369-493-44624	Sequence 44624, A
C	3	95	20.8	822900	16	US-10-292-798-1393	Sequence 1393, Ap
C	4	94.5	20.7	6395	17	US-10-437-963-34802	Sequence 34802, A
	5	94.5	20.7	24081	15	US-10-132-134-13	Sequence 13, Appli
	6	94.5	20.7	52101	15	US-10-132-134-1	Sequence 1, Appli
C	7	94	20.6	699	14	US-10-006-522-41	Sequence 41, Appli
	8	93.5	20.5	943	13	US-10-245-752-111	Sequence 111, App
	9	93.5	20.5	943	13	US-10-245-859-111	Sequence 111, App
	10	93.5	20.5	943	15	US-10-245-103-111	Sequence 111, App
	11	93.5	20.5	943	15	US-10-245-107-111	Sequence 111, App
	12	93.5	20.5	943	15	US-10-245-143-111	Sequence 111, App
	13	93.5	20.5	943	15	US-10-245-771-111	Sequence 111, App
	14	93.5	20.5	943	15	US-10-245-851-111	Sequence 111, App
	15	93.5	20.5	943	15	US-10-245-883-111	Sequence 111, App
	16	93.5	20.5	943	15	US-10-237-535-111	Sequence 111, App
	17	93.5	20.5	943	15	US-10-238-183-111	Sequence 111, App
	18	93.5	20.5	943	15	US-10-238-283-111	Sequence 111, App
	19	93.5	20.5	943	15	US-10-238-370-111	Sequence 111, App
	20	93.5	20.5	943	15	US-10-245-055-111	Sequence 111, App
	21	93.5	20.5	943	15	US-10-245-447-111	Sequence 111, App
	22	93.5	20.5	943	15	US-10-245-730-111	Sequence 111, App
	23	93.5	20.5	943	15	US-10-245-739-111	Sequence 111, App
	24	93.5	20.5	943	15	US-10-246-210-111	Sequence 111, App
	25	93.5	20.5	943	15	US-10-239-196-111	Sequence 111, App
	26	93.5	20.5	943	15	US-10-243-024-111	Sequence 111, App
	27	93.5	20.5	943	15	US-10-243-409-111	Sequence 111, App
	28	93.5	20.5	943	15	US-10-245-621-111	Sequence 111, App
	29	93.5	20.5	943	15	US-10-245-880-111	Sequence 111, App
	30	93.5	20.5	943	15	US-10-245-033-111	Sequence 111, App
	31	93.5	20.5	943	15	US-10-243-095-111	Sequence 111, App
	32	93.5	20.5	943	15	US-10-245-185-111	Sequence 111, App
	33	93.5	20.5	943	15	US-10-245-427-111	Sequence 111, App
	34	93.5	20.5	943	15	US-10-245-473-111	Sequence 111, App
	35	93.5	20.5	943	15	US-10-245-770-111	Sequence 111, App
	36	93.5	20.5	943	15	US-10-245-877-111	Sequence 111, App
	37	93.5	20.5	943	15	US-10-246-976-111	Sequence 111, App
	38	93.5	20.5	943	15	US-10-243-320-111	Sequence 111, App
	39	93.5	20.5	943	15	US-10-242-743-111	Sequence 111, App
	40	93.5	20.5	943	15	US-10-242-845-111	Sequence 111, App
	41	93.5	20.5	943	15	US-10-237-636-111	Sequence 111, App
	42	93.5	20.5	943	15	US-10-238-325-111	Sequence 111, App
	43	93.5	20.5	943	15	US-10-238-346-111	Sequence 111, App
	44	93.5	20.5	943	15	US-10-238-411-111	Sequence 111, App
	45	93.5	20.5	943	15	US-10-243-124-111	Sequence 111, App

## ALIGNMENTS

RESULT 1  
US-09-972-032-1  
; Sequence 1, Application US/0972032  
; Publication No. US20020086361A1  
; GENERAL INFORMATION:  
; APPLICANT: Case Western Reserve University  
; APPLICANT: Montano, Monica  
; APPLICANT: Sutton, Amelia  
; TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology  
; FILE REFERENCE: 27708/04003  
; CURRENT APPLICATION NUMBER: US/09/972,032  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/238,190  
; PRIOR FILING DATE: 2000-10-05  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 990  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-972-032-1

Alignment Scores:



```

;
; NAME/KEY: modified_base
; LOCATION: (67605)..(67704)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (74625)..(74724)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (85854)..(85953)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367573)..(367573)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367588)..(367588)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367595)..(367595)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367613)..(367614)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367985)..(367985)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367993)..(367993)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367995)..(367996)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367998)..(367998)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370273)..(370277)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370279)..(370279)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370281)..(370282)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370285)..(370287)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370289)..(370291)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (385643)..(385742)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (409961)..(410060)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:

```

```

; NAME/KEY: modified_base
; LOCATION: (410096)..(410096)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (417384)..(417483)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (724960)..(725059)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (726106)..(726205)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (727470)..(727569)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (728849)..(728948)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (730296)..(730395)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (731863)..(731962)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (734124)..(734223)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (734441)..(734441)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (735752)..(735851)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (738576)..(738675)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (740924)..(741023)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (744360)..(744459)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (748430)..(748529)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (754323)..(754422)
; OTHER INFORMATION: a, t, c, g, unknown or other
;
;
; Alignment Scores:
; Pred. No.: 348
; Length: 822900
; Score: 95.00
; Matches: 37
; Percent Similarity: 37.39%
; Conservative: 6
; Best Local Similarity: 32.17%
; Mismatches: 29
; Query Match: 20.83%
; Indels: 43
; DB: 16
; Gaps: 5
;
; US-09-972-032-2 (1-79) x US-10-292-798-1393 (1-822900)

```



```

; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

Alignment Scores:
Pred. No.: 34.2 Length: 52101
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 15 Gaps: 3

US-09-972-032-2 (1-79) x US-10-132-134-1 (1-52101)

Qy 9 SexAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-GI 28
Db 30123 TCGGGCGCCCTCGAGGACGCGCAGCGCCGCGCCCTGCGCGAGGCGCTGGACTGC 30182

Qy 28 uGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48
Db 30183 CGGGTCAAGGTCGCTCAACTGGGGCTACTGGGGCAACGCTCCCGCAGCAGCTCTCTGCGCGAC 30242

Qy 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
Db 30243 GTGAGTCCATGGAGTGGCCCGCGATCGCCCGCGCAGCGCGAGTGGCGCACTGGAGCGC 30302

Qy 61 aGlyLeuThrValArg 66
Db 30303 CTCCTGCGCGCGCGC 30318

RESULT 7
US-10-006-922-41/c
; Sequence 41, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Pradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-006-922-41

Alignment Scores:
Pred. No.: 0.856 Length: 699
Score: 94.00 Matches: 33
Percent Similarity: 43.02% Conservative: 4
Best Local Similarity: 38.37% Mismatches: 31
Query Match: 20.61% Indels: 18
DB: 14 Gaps: 6

US-09-972-032-2 (1-79) x US-10-006-922-41 (1-699)

; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

Alignment Scores:
Pred. No.: 34.2 Length: 52101
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 15 Gaps: 3

US-09-972-032-2 (1-79) x US-10-132-134-1 (1-52101)

Qy 9 SexAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-GI 28
Db 30123 TCGGGCGCCCTCGAGGACGCGCAGCGCCGCGCCCTGCGCGAGGCGCTGGACTGC 30182

Qy 28 uGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48
Db 30183 CGGGTCAAGGTCGCTCAACTGGGGCTACTGGGGCAACGCTCCCGCAGCAGCTCTCTGCGCGAC 30242

Qy 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
Db 30243 GTGAGTCCATGGAGTGGCCCGCGATCGCCCGCGCAGCGCGAGTGGCGCACTGGAGCGC 30302

Qy 61 aGlyLeuThrValArg 66
Db 30303 CTCCTGCGCGCGCGC 30318

RESULT 8
US-10-245-752-111
; Sequence 111, Application US/10245752
; Publication No. US20030064473A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC66
; CURRENT APPLICATION NUMBER: US/10/245,752
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-752-111

Alignment Scores:
Pred. No.: 1.25 Length: 943

```

Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7  
 Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 13 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-752-111 (1-943)

QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaSerAlaHisTrpThrGly 21  
 Db 138 TGTCCAGACGCTGGAGATGGCGGCGAGCGGCGGCTGTGGAGTCTCTGCTGGC 137  
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyLeuGlyProGlu----- 37  
 Db 198 TGGTGGACAGGACCGGGAGGCGGA-----GCCCTGGGGCCAGAGCGGCGAGGTGG 251  
 QY 38 -----GlyGlnAlaSerProThrPro 44  
 Db 252 ACGCACATGACTGTGAGCGCTGGGCTCCGAGGAGCGGCTTCAGAGTCCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGCAGCGTCAAACTGCAGTTCGACATGATGCGGCTCGAACCTGGCCACGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGCTCAGCGGAGGCGAGTCACTTCTCTGGGGCTGGTGGCTGCCCCCTGCTGTAC 431  
 QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76  
 Db 432 CCGACGCGCCCGTGTCTGGGAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491  
 QY 76 tGly 77  
 Db 492 TGGT 495

# RESULT 9

US-10-245-859-111  
 ; Sequence 111, Application US/10245859  
 ; Publication No. US20030064474A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin  
 ; APPLICANT: Eaton, Dan  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Grimaldi, J. Christopher  
 ; APPLICANT: Gurney, Austin  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stephan, Jean-Phillippe  
 ; APPLICANT: Watanabe, Colin  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; APPLICANT: Fong, Sherman  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3630R1C78  
 ; CURRENT APPLICATION NUMBER: US/10/245, 859  
 ; CURRENT FILING DATE: 2002-09-16  
 ; PRIOR APPLICATION NUMBER: 10/197942  
 ; PRIOR FILING DATE: 2002-07-18  
 ; PRIOR APPLICATION NUMBER: 60/059114  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/063046  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/065027  
 ; PRIOR FILING DATE: 1997-11-10  
 ; PRIOR APPLICATION NUMBER: 60/079689  
 ; PRIOR FILING DATE: 1998-03-27  
 ; PRIOR APPLICATION NUMBER: 60/086478  
 ; PRIOR FILING DATE: 1998-05-22  
 ; PRIOR APPLICATION NUMBER: 60/087607  
 ; PRIOR FILING DATE: 1998-06-02  
 ; PRIOR APPLICATION NUMBER: 60/089801

; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: 60/090557  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090689  
 ; PRIOR FILING DATE: 1998-06-25  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 116  
 ; SEQ ID NO 111  
 ; LENGTH: 943  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; US-10-245-859-111

Alignment Scores:  
 Pred. No.: 1.25 Length: 943  
 Score: 93.50 Matches: 36  
 Percent Similarity: 35.25% Conservative: 7  
 Best Local Similarity: 29.51% Mismatches: 31  
 Query Match: 20.50% Indels: 48  
 DB: 13 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-859-111 (1-943)

QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaSerAlaHisTrpThrGly 21  
 Db 138 TGTCCAGACGCTGGAGATGGCGGCGAGCGGCGGCTGTGGAGTCTCTGCTGGC 197  
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyLeuGlyProGlu----- 37  
 Db 198 TGGTGGACAGGACCGGGAGGCGGA-----GCCCTGGGGCCAGAGCGGCGAGGTGG 251  
 QY 38 -----GlyGlnAlaSerProThrPro 44  
 Db 252 ACGCACATGACTGTGAGCGCTGGGCTCCGAGGAGCGGCTTCAGAGTCCC 311  
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52  
 Db 312 GAGCAGCGTCAAACTGCAGTTCGACATGATGCGGCTCGAACCTGGTGGCCACGCGCG 371  
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66  
 Db 372 CGCTCAGCGGAGGCGAGTCACTTCTCTGGGGCTGGTGGCTGCCCCCTGCTGTAC 431  
 QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76  
 Db 432 CCGACGCGCCCGTGTCTGGGAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491  
 QY 76 tGly 77  
 Db 492 TGGT 495

# RESULT 10

US-10-245-103-111  
 ; Sequence 111, Application US/10245103  
 ; Publication No. US20030068778A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin  
 ; APPLICANT: Eaton, Dan  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Grimaldi, J. Christopher  
 ; APPLICANT: Gurney, Austin  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stephan, Jean-Phillippe  
 ; APPLICANT: Watanabe, Colin  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; APPLICANT: Fong, Sherman  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3630R1C112  
 ; CURRENT APPLICATION NUMBER: US/10/245, 103  
 ; CURRENT FILING DATE: 2002-09-17



```

; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-103-111
Alignment Scores:
Pred. No.: 1.25          Length: 943
Score: 93.50           Matches: 36
Percent Similarity: 35.25%      Conservative: 7
Best Local Similarity: 29.51%   Mismatches: 31
Query Match: 20.50%           Indels: 48
DB: 15                   Gaps: 6
US-09-972-032-2 (1-79) x US-10-245-103-111 (1-943)
QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCTGGGGCTGTGGAGTCTCTGCTGC 197
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyIleGlyProGlu----- 37
Db 198 TGTGGACAGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGCGCGCTTCACAGGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCCAGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGCTCACCAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 431
QY 66 gApArgProGlnLeu-----GlyGlnLeuCysMe 76
Db 432 CCGACGCCCGTGTCTGGGAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495
RESULT 11
US-10-245-107-111
; Sequence 111, Application US/10245107
; Publication No. US20030068779A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan

```

```

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC71
; CURRENT APPLICATION NUMBER: US/10/245,107
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-107-111
Alignment Scores:
Pred. No.: 1.25          Length: 943
Score: 93.50           Matches: 36
Percent Similarity: 35.25%      Conservative: 7
Best Local Similarity: 29.51%   Mismatches: 31
Query Match: 20.50%           Indels: 48
DB: 15                   Gaps: 6
US-09-972-032-2 (1-79) x US-10-245-107-111 (1-943)
QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCTGGGGCTGTGGAGTCTCTGCTGC 197
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyIleGlyProGlu----- 37
Db 198 TGTGGACAGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGCGCGCTTCACAGGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCCAGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGCTCACCAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 431

```



## Alignment Scores:

```

Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-771-111 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTCCAGACGCTGGAGATGGGCGAGCGCAGCGCTGGGGCTGTGGAGTCTGTGGC 197
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 37
Db 198 TGTGTGACAGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGCGCGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGAGCGCGGCTTCCAGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCGCACGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66
Db 372 CGTCAACGAGCGACGCTACCTTCCTCTGGGGCTGGTGGCGCTGCCCTGTCTCAC 431
QY 66 GAsPArgProGlnLeu-----GlyGluLeuCysMe 76
Db 432 CCGAGCGCCCGTGTGGAGGAGGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495

```

## RESULT 14

```

US-10-245-851-111
; Sequence 111, Application US/10245851
; Publication No. US20030068782A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630R1C93
; CURRENT APPLICATION NUMBER: US/10/245,851
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607

```

```

; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-851-111

```

## Alignment Scores:

```

Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-851-111 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTCCAGACGCTGGAGATGGGCGAGCGCAGCGCTGGGGCTGTGGAGTCTGTGGC 197
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 37
Db 198 TGTGTGACAGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGCGCGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGAGCGCGGCTTCCAGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCGCACGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66
Db 372 CGTCAACGAGCGACGCTACCTTCCTCTGGGGCTGGTGGCGCTGCCCTGTCTCAC 431
QY 66 GAsPArgProGlnLeu-----GlyGluLeuCysMe 76
Db 432 CCGAGCGCCCGTGTGGAGGAGGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495

```

## RESULT 15

```

US-10-245-883-111
; Sequence 111, Application US/10245883
; Publication No. US20030068783A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630R1C70

```

```
; CURRENT APPLICATION NUMBER: US/10/245,883
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-883-111

Alignment Scores:
Pred. No.:      1.25      Length:      943
Score:          93.50     Matches:     36
Percent Similarity: 35.25%  Conservative: 7
Best Local Similarity: 29.51% Mismatches:    31
Query Match:      20.50%  Indels:      48
DB:               15      Gaps:         6

US-09-972-032-2 (1-79) x US-10-245-883-111 (1-943)

Qy      2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db      138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCGTGGGGCTGTGGAGTCTCTGCTGC 197
Qy      22 LeuTrpThrGlyLeuGlyGluGlyGluGlyGlyGlyProGlu----- 37
Db      198 TGTGGACAGACCCGGGAGGCCGA-----GCCCTGGGGCCAGAGCGCGCCAGGTGG 251
Qy      38 -----GlyGlnAlaSerProThrPro 44
Db      252 ACGCACATGACTGTAGGCGCTGGGCTGGGCTCCGAGGCGAGCGGCTCCAGGAGTCCG 311
Qy      45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db      312 GAGGCACCGTCAAACTGCAGTTGCATGATGCGCGCTGCACCTGTGGCCCGCCGCGCG 371
Qy      53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db      372 CGCTCACCAGGCGCAGCTCACCTTCCTCTCTGGGGCTGTGGGCTGCCCCCTGCTGTAC 431
Qy      66 GAspArgProGlnLeu-----GlyGlnLeuCysMe 76
Db      432 CCGACGCCCCGCTGCTGGGAGAGGCCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
Qy      76 tGly 77
Db      492 TGGT 495
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 10, 2004, 16:42:00 ; Search time 71 seconds  
(without alignments)  
617.481 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRRVSGAGCFADAHWT.....SAGLTVRDPLGELCMGRG 79

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0972032/runat\_06072004\_121453\_8423/app\_query.fasta\_1.263  
-DB=Issued Patents NA -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0972032@cgn 1.1.69 @runat\_06072004\_121453\_8423 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	20.4	558	4	US-09-252-991A-5885
2	91	20.0	2043	4	US-09-614-912-171
3	88	19.3	1386	4	US-09-433-248A-1
4	88	19.3	3625	4	US-09-023-655-1180
5	87.5	19.2	497	4	US-09-621-976-12967
6	86.5	19.0	603	4	US-09-252-991A-5342
7	86.5	19.0	909	4	US-09-252-991A-5304
8	86.5	19.0	981	4	US-09-252-991A-5231
9	86.5	19.0	2592	4	US-09-252-991A-5260
10	86.5	19.0	4403765	3	US-09-103-840A-2
11	86.5	19.0	4411529	3	US-09-103-840A-1
12	86	18.9	23673	4	US-09-773-816-1

13	85	18.6	1758	3	US-08-753-247-13	Sequence 13, Appl
14	85	18.6	1776	3	US-08-753-247-17	Sequence 17, Appl
15	85	18.6	1794	3	US-08-753-247-20	Sequence 20, Appl
16	85	18.6	2130	3	US-08-753-247-5	Sequence 5, Appl
17	85	18.6	2142	3	US-08-753-247-8	Sequence 8, Appl
18	85	18.6	2151	1	US-08-477-254A-5	Sequence 5, Appl
19	85	18.6	2151	2	US-08-477-254B-5	Sequence 5, Appl
20	85	18.6	2151	2	US-08-428-734B-5	Sequence 5, Appl
21	85	18.6	2151	3	US-08-713-556F-5	Sequence 5, Appl
22	85	18.6	2160	3	US-08-753-247-11	Sequence 11, Appl
23	85	18.6	2385	1	US-07-885-972A-1	Sequence 1, Appl
24	85	18.6	2385	2	US-08-745-880-1	Sequence 1, Appl
25	85	18.6	2385	2	US-08-480-382-1	Sequence 1, Appl
26	85	18.6	2388	4	US-10-133-910-1	Sequence 3, Appl
27	85	18.6	4405	1	US-07-885-972A-3	Sequence 3, Appl
28	85	18.6	4405	2	US-08-745-880-3	Sequence 3, Appl
29	85	18.6	4405	2	US-08-480-382-3	Sequence 3, Appl
30	84.5	18.5	463	4	US-09-841-334A-7	Sequence 7, Appl
31	84.5	18.5	777	4	US-09-252-991A-13091	Sequence 13091, A
32	84.5	18.5	1278	4	US-09-252-991A-9269	Sequence 9269, Ap
33	84.5	18.5	1362	4	US-09-252-991A-9298	Sequence 9298, Ap
34	84.5	18.5	1521	4	US-09-252-991A-9281	Sequence 9281, Ap
35	84.5	18.5	1632	4	US-09-252-991A-13259	Sequence 13259, A
36	83	18.2	477	4	US-09-252-991A-7999	Sequence 7999, Ap
37	83	18.2	813	4	US-09-252-991A-8201	Sequence 8201, Ap
38	83	18.2	972	4	US-09-252-991A-8102	Sequence 8102, Ap
39	83	18.2	1236	3	US-09-105-343A-1	Sequence 1, Appl
40	83	18.2	1919	4	US-09-614-912-175	Sequence 175, App
41	83	18.2	1944	4	US-09-252-991A-8062	Sequence 8062, Ap
42	83	18.2	7000	4	US-09-851-896-3	Sequence 3, Appl
43	82.5	18.1	2275	4	US-08-850-977-1	Sequence 1, Appl
44	82.5	18.1	30001	1	US-08-125-468-1	Sequence 1, Appl
45	82.5	18.1	30001	2	US-08-474-933-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-5885

Sequence 5885, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 5885

LENGTH: 558

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5885

Alignment Scores:				
Pred. No.:	2.52	Length:	558	
Score:	93.00	Matches:	28	
Percent Similarity:	28.30%	Conservative:	8	
Best Local Similarity:	29.79%	Mismatches:	26	
Query Match:	20.39%	Indels:	32	
DB:	4	Gaps:	4	

US-09-972-032-2 (1-79) x US-09-252-991A-5885 (1-558)

Qy 4 ArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrp 23

Db 207 CGTCCTCGACGTCGCCGCGAGGATGATTCCGCGACACGTTTC----- 254



ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,655

FILING DATE: HERewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen J.

REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA-0001 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 1180:

SEQUENCE CHARACTERISTICS:

LENGTH: 3625 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GENBANK

CLONE: G189177

US-09-023-655-1180

#### Alignment Scores:

Pred. No.: 57.7 Length: 3625  
Score: 88.00 Matches: 32  
Percent Similarity: 39.33% Conservative: 3  
Best Local Similarity: 35.96% Mismatches: 30  
Query Match: 19.30% Indels: 24  
DB: 4 Gaps: 5

US-09-972-032-2 (1-79) x US-09-023-655-1180 (1-3625)

Qy 4 ArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrp 23  
Db 312 AGAAGAGCGCGCGCGAGCGCGGAGCCCGG-----GCGGCGCCGAGT 271  
Qy 24 ThrGlyLeuGlyGluGlyGlnGlyGlyLeuGlyGlyProGlyGlnAlaSerProThr 43  
Db 270 GCGGGTCGGGCGAGCGGAGCGGCGGCGGCGGCGGCGCA--GGCCAGTCAGCGCGCGC 214  
Qy 44 ProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpPro-----58  
Db 213 CATGCGCAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154  
Qy 59 ---Trp-----SerAlaGlyLeuThrValArgAspArgProGln 70  
Db 153 TACTGGCGGCGGCGTCCGCGCTCGGCGCGCTCGGCGTCCGCGGAGCGGCGCGG-----100  
Qy 71 LeuGlyGluLeuCysMetGlyArgGly 79  
Db 99 -----TGGCTGCGTGCAGGCGCGAGGG 79

#### RESULT 5

US-09-621-976-12967/c

; Sequence 12967, Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTS and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 12967

; LENGTH: 497

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-621-976-12967

#### Alignment Scores:

Pred. No.: 6.87 Length: 497  
Score: 87.50 Matches: 21  
Percent Similarity: 48.08% Conservative: 4  
Best Local Similarity: 40.38% Mismatches: 15  
Query Match: 19.19% Indels: 12  
DB: 4 Gaps: 3

US-09-972-032-2 (1-79) x US-09-621-976-12967 (1-497)

Qy 19 TrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGlyLeuGlyProGluGly 38  
Db 174 TGGCGGGGTGTTGGCCA---CTAGGTGGTGGAGGTAGGATATATGGGTCCTTCA 118  
Qy 39 GlnAlaSerProThrPro-----AspCysAlaSerArgTrp-Pro-----51  
Db 117 CAATCTGGGCTCTTCCTTATTAGGTGCAGGGTGTTCAGAGATGTCCTCGGCTCC 58  
Qy 52 -----ArgSerAlaSerArgTrpProTrp 59  
Db 57 TCAGGCTGGACGTCGCGGCGAGCAGGTCGCGCTGG 24

#### RESULT 6

US-09-252-991A-5342/c

; Sequence 5342, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5342

; LENGTH: 603

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5342

#### Alignment Scores:

Pred. No.: 10.5 Length: 603  
Score: 86.50 Matches: 24  
Percent Similarity: 47.33% Conservative: 3  
Best Local Similarity: 42.11% Mismatches: 27  
Query Match: 18.97% Indels: 3  
DB: 4 Gaps: 2

US-09-972-032-2 (1-79) x US-09-252-991A-5342 (1-603)

Qy 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThrGly 21  
Db 174 CGAAACCGTCGACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 115  
Qy 22 LeuTrpThrGlyLeuGlyGlnGlyGlyLeuGlyProGluGlyGlnAlaSer 41

Db 114 ATCGGACAGGAGCTCTCGCAATGGCGCGCAAGCAGCCCTGTT---CGAGCAGGC 58  
 QY 42 ProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPPro 58  
 Db 57 CGTCGACCAAGTGGCGCAGCAGCGATGGCCAGCCCTGCGCGGCGATGGCGG 7

## RESULT 7

US-09-252-991A-5304/c  
 ; Sequence 5304, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 5304  
 ; LENGTH: 909  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-5304

## Alignment Scores:

Pred. No.: 16.6 Length: 909  
 Score: 86.50 Matches: 24  
 Percent Similarity: 47.37% Conservative: 3  
 Best Local Similarity: 42.11% Mismatches: 27  
 Query Match: 18.97% Indels: 3  
 DB: 4 Gaps: 2

US-09-972-032-2 (1-79) x US-09-252-991A-5304 (1-909)

QY 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 868 CGAAACCGTCGACGAGCGCTGGCGCGCTTGGCTCGCGCAATACCGCATAGTACCCGT 809  
 QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 41  
 Db 808 ATCGGACAGGAGCTCTCGCAATGGCGCGCAAGCAGCCCTGTT---CGAGCAGGC 752  
 QY 42 ProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPPro 58  
 Db 751 CGTCGACCAAGTGGCGCAGCAGCGATGGCCAGCCCTGCGCGGCGATGGCGG 701

## RESULT 8

US-09-252-991A-5231  
 ; Sequence 5231, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 5231  
 ; LENGTH: 981  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-5231

## Alignment Scores:

Pred. No.: 18.1 Length: 981  
 Score: 86.50 Matches: 24  
 Percent Similarity: 47.37% Conservative: 3  
 Best Local Similarity: 42.11% Mismatches: 27  
 Query Match: 18.97% Indels: 3  
 DB: 4 Gaps: 2

US-09-972-032-2 (1-79) x US-09-252-991A-5231 (1-981)

QY 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 781 CGAAACCGTCGACGAGCGCTGGCGCGCTTGGCTCGCGCAATACCGCATAGTACCCGT 840  
 QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 41  
 Db 841 ATCGGACAGGAGCTCTCGCAATGGCGCGCAAGCAGCCCTGTT---CGAGCAGGC 897  
 QY 42 ProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPPro 58  
 Db 898 CGTCGACCAAGTGGCGCAGCAGCGATGGCCAGCCCTGCGCGGCGATGGCGG 948

## RESULT 9

US-09-252-991A-5260  
 ; Sequence 5260, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 5260  
 ; LENGTH: 2592  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-5260

## Alignment Scores:

Pred. No.: 53.9 Length: 2592  
 Score: 86.50 Matches: 24  
 Percent Similarity: 47.37% Conservative: 3  
 Best Local Similarity: 42.11% Mismatches: 27  
 Query Match: 18.97% Indels: 3  
 DB: 4 Gaps: 2

US-09-972-032-2 (1-79) x US-09-252-991A-5260 (1-2592)

QY 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThrGly 21  
 Db 93 CGAAACCGTCGACGAGCGCTGGCGCGCTTGGCTCGCGCAATACCGCATAGTACCCGT 152  
 QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 41  
 Db 153 ATCGGACAGGAGCTCTCGCAATGGCGCGCAAGCAGCCCTGTT---CGAGCAGGC 209  
 QY 42 ProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPPro 58  
 Db 210 CGTCGACCAAGTGGCGCAGCAGCGATGGCCAGCCCTGCGCGGCGATGGCGG 260

## RESULT 10

US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.



```

; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 2,11e+05 Length: 4403765
Score: 86.50 Matches: 28
Percent Similarity: 39.33% Conservative: 7
Best Local Similarity: 31.46% Mismatches: 20
Query Match: 18.97% Indels: 35
DB: 3 Gaps: 6

US-09-972-032-2 (1-79) x US-09-103-840A-2 (1-4403765)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 2196902 TGGCGTCGACCGGTGG-ATT-:::-----TGCAGTTTCGCG 2196933
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 41
Db 2196934 TTGTGG---GGCACCAGTGTGGTAGCGACGTCGCGATTGGG-----GGTTCGTGGAAG 2196984
QY 42 ProThr-----:::----- 43
Db 2196985 CCGACGGTGTAAAGGGTGGCGTAGCGGTGGCTATTTCGGCGGCACCTTCGTTCGACT 2197044
QY 44 ---ProAspCysAlaSerArgTrp-----ProArgSerAlaSerArgTrpProTrp 59
Db 2197045 CCGATGAGTTGTTCGCGCGTGGCGGATATCCCGGCCCAATGCTTGTGTGGCGCTGG 2197104
QY 60 SerAlaGlyLeuThrValArgAspArg 68
Db 2197105 CGAGCTGCTTGTTCGCGAAGGTGCGG 2197131

RESULT 11
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 2,11e+05 Length: 4411529
Score: 86.50 Matches: 28
Percent Similarity: 39.33% Conservative: 7
Best Local Similarity: 31.46% Mismatches: 20
Query Match: 18.97% Indels: 35
DB: 3 Gaps: 6

US-09-972-032-2 (1-79) x US-09-103-840A-1 (1-4411529)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 2199603 TGGCGTCGACCGGTGG-ATT-:::-----TGCAGTTTCGCG 2199634
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 41
Db 2199635 TTGTGG---GGCACCAGTGTGGTAGCGACGTCGCGATTGGG-----GGTTCGTGGAAG 2199685
QY 42 ProThr-----:::----- 43
Db 2199686 CCGACGGTGTAAAGGGTGGCGTAGCGGTGGCTATTTCGGCGGCACCTTCGTTCGACT 2199745
QY 44 ---ProAspCysAlaSerArgTrp-----ProArgSerAlaSerArgTrpProTrp 59
Db 2199746 CCGATGAGTTGTTCGCGCGTGGCGGATATCCCGGCCCAATGCTTGTGTGGCGCTGG 2199805
QY 60 SerAlaGlyLeuThrValArgAspArg 68
Db 2199806 CGAGCTGCTTGTTCGCGAAGGTGCGG 2199832

RESULT 12
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; TITLE OF INVENTION: ANTAGONISTS
; FILE REFERENCE: 28600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23623)
; OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Alignment Scores:
Pred. No.: 717 Length: 23673
Score: 86.00 Matches: 23
Percent Similarity: 41.27% Conservative: 3
Best Local Similarity: 36.51% Mismatches: 23
Query Match: 18.86% Indels: 14
DB: 4 Gaps: 3

US-09-972-032-2 (1-79) x US-09-773-816-1 (1-23673)

QY 6 ArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGly 25
Db 18526 CGCAGAACCTCGCGCGCGTGGCGGACCTGGAGTTCCTCGCGGAGGTGGTGGACCGGC 18585
QY 26 LeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 41
Db 26 LeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 41

```

Db 18586 TCGGCTCGGCGCTGTCGGCACCGCCCGCGAGTTGTCGGCGGCGCAGCAGC 18645

QY 42 ProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAla 61  
 Db 18646 -----GGGTCCGGTGG-----CGCGCGCCCTGGTCAGCA 18675

QY 62 GlyLeuThr 64  
 Db 18676 GGCCGGACA 18684

RESULT 13

US-08-753-247-13  
 ; Sequence 13, Application US/08753247  
 ; Patent No. 6210929  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SCHLOKAT, Uwe  
 ; APPLICANT: FISCHER, Bernhard  
 ; APPLICANT: FALKNER, Falko-Guenther  
 ; APPLICANT: DORNER, Friedrich  
 ; APPLICANT: BIBL, Johann  
 ; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN  
 ; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A  
 ; TITLE OF INVENTION: HETEROLOGOUS SEQUENCE  
 ; NUMBER OF SEQUENCES: 29  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/08/753,247  
 ; APPLICATION NUMBER: 29  
 ; FILING DATE: 22-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: AT 1928/95  
 ; FILING DATE: 24-NOV-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 40433/149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1758 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1755  
 ; FEATURE:  
 ; NAME/KEY: mat\_peptide  
 ; LOCATION: 1..1755  
 ; US-08-753-247-13

Alignment Scores:  
 Pred. No.: 47.5 Length: 1758  
 Score: 85.00 Matches: 30  
 Percent Similarity: 34.69% Conservative: 4  
 Best Local Similarity: 30.61% Mismatches: 34  
 Query Match: 18.64% Indels: 30  
 DB: 3 Gaps: 6

US-09-972-032-2 (1-79) x US-08-753-247-13 (1-1758)

QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22  
 Db 615 GGCCAAACAACGGTGTCTGTGGTGTAGGTGTGGCTTACCAACGCCCGCATTTGGAGGGTGGC 674

QY 23 TrpThrGly----LeuGlyGluGlyClnGluGlyGly-----IleGlyProGluGlyGln 39  
 Db 675 CATGCTGATGCGGAGGTTCACAGATGCAGTGCAGGACACGCTCGTGGGCTTGAACCCCAA 734

QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50  
 Db 735 CCACATCCACATCTACAGTGCAGCTGGGGCCCCGAGGATGACGCGAAGACAGTGGATGG 794

QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61  
 Db 795 GCCAGCCCGCTCGCCGAGGAGGCTTCTTCCGTGGGTTAGCCAGGCGCCGAGGGGGCT 854

QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79  
 Db 855 GGGCTC-----CATCTTTGCTGGGCTCGGG 881

RESULT 14

US-08-753-247-17  
 ; Sequence 17, Application US/08753247  
 ; Patent No. 6210929  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SCHLOKAT, Uwe  
 ; APPLICANT: FISCHER, Bernhard  
 ; APPLICANT: FALKNER, Falko-Guenther  
 ; APPLICANT: DORNER, Friedrich  
 ; APPLICANT: BIBL, Johann  
 ; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN  
 ; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A  
 ; TITLE OF INVENTION: HETEROLOGOUS SEQUENCE  
 ; NUMBER OF SEQUENCES: 29  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/08/753,247  
 ; APPLICATION NUMBER: 29  
 ; FILING DATE: 22-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: AT 1928/95  
 ; FILING DATE: 24-NOV-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 40433/149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1776 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS

```
; LOCATION: 1..1773
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..1773
US-08-753-247-17

Alignment Scores:
Pred. No.: 48      Length: 1776
Score: 85.00      Matches: 30
Percent Similarity: 34.69%      Conservative: 4
Best Local Similarity: 30.61%      Mismatches: 34
Query Match: 18.64%      Indels: 30
DB: 3      Gaps: 6

US-09-972-032-2 (1-79) x US-08-753-247-17 (1-1776)
QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
Db 615 GGCCAAACACGGTCTGTGGTGTAGTGTGGCTTACAGCCCGCATTTGGAGGGGTGG 674
QY 23 TrpThrGly---LeuGlyGluGlyGlnGluGlyGly-----IleGlyProGluGlyGln 39
Db 675 CATGCTGGATGGCGAGGTGCACAGATGCAGTGGAGGACGGTCTGGCTGAACCCCAA 734
QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50
Db 735 CCACATCCACATCTACAGTGCACAGTGGGCGCCGAGGATGACGCCAAGACAGTGGATGG 794
QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61
Db 795 GCACGCCCGCTGCCGAGGAGGCGCTTCTTCGTTGGGTAGCCAGGCGCGAGGGGGCT 854
QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 855 GGGCTC-----CATCTTTGTCTGGGCTCGGG 881

RESULT 15
US-08-753-247-20
; Sequence 20, Application US/08753247
; Patent No. 6210929
; GENERAL INFORMATION:
; APPLICANT: SCHLOKAT, Uwe
; APPLICANT: FISCHER, Bernhard
; APPLICANT: FALKNER, Falko-Guenther
; APPLICANT: DORNER, Friedrich
; APPLICANT: EIBL, Johann
; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN
; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A
; TITLE OF INVENTION: HETEROLOGOUS SEQUENCE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,247
; FILING DATE: 22-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AT 1928/95
; FILING DATE: 24-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40433/149
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1791
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..1791
US-08-753-247-20

Alignment Scores:
Pred. No.: 48.6      Length: 1794
Score: 85.00      Matches: 30
Percent Similarity: 34.69%      Conservative: 4
Best Local Similarity: 30.61%      Mismatches: 34
Query Match: 18.64%      Indels: 30
DB: 3      Gaps: 6

US-09-972-032-2 (1-79) x US-08-753-247-20 (1-1794)
QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
Db 615 GGCCAAACACGGTCTGTGGTGTAGTGTGGCTTACAGCCCGCATTTGGAGGGGTGG 674
QY 23 TrpThrGly---LeuGlyGluGlyGlnGluGlyGly-----IleGlyProGluGlyGln 39
Db 675 CATGCTGGATGGCGAGGTGCACAGATGCAGTGGAGGACGGTCTGGCTGAACCCCAA 734
QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50
Db 735 CCACATCCACATCTACAGTGCACAGTGGGCGCCGAGGATGACGCCAAGACAGTGGATGG 794
QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61
Db 795 GCACGCCCGCTGCCGAGGAGGCGCTTCTTCGTTGGGTAGCCAGGCGCGAGGGGGCT 854
QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 855 GGGCTC-----CATCTTTGTCTGGGCTCGGG 881
```

Search completed: July 10, 2004, 18:58:39  
Job time : 1527 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 10, 2004, 16:36:45 ; Search time 2885 Seconds  
(without alignments)  
817.717 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 456  
Sequence: 1 MCGRRPRVSAGCGFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=cg2 1/USPTO.spool/US09972032/runat\_06072004\_121452\_8401/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -STAR=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972032@cgn 1.1 3437 @runat\_06072004\_121452\_8401 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmb:\*  
5: em\_estov:\*  
6: em\_estro:\*  
7: em\_etc:\*  
8: em\_etc1:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_etc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	1467	11	BC039500 Homo sapi
2	421	92.3	1201	13	EX406138 BX406138
3	395	86.6	760	12	BM980506 UI-CF-EN1
4	274	60.1	691	13	BU620427 UI-H-FL1
5	252	55.3	694	13	BU684594 UI-CF-EN1
6	249.5	54.7	1084	12	BM806108 AGENCOURT
7	238	52.2	669	12	BM679577 UI-E-BOO
8	187	41.0	632	13	BU617331 UI-H-DF0
9	187	41.0	633	14	CA313142 UI-CF-FNO
10	183	40.1	623	14	CD742581 UI-H-FT2
11	166	36.4	628	13	BQ447041 UI-H-EUL
12	98.5	21.6	486	12	BM484011 537701 MA
13	96.5	21.2	524	29	CC514259 CH240_359
14	96.5	21.2	1362	13	BQ644396 AGENCOURT
15	95.5	20.9	872	12	BG539339 602567951
16	95.5	20.9	906	14	CD171626 AGENCOURT
17	95.5	20.9	1443	28	BZ575371 meh2_4430
18	95	20.8	563	9	AW047982 UI-N-BH1
19	94.5	20.7	378	13	BU997672 H108K20r
20	94.5	20.7	557	9	AV665399 AV665399
21	94.5	20.7	632	14	CA756603 BR0500020
22	94.5	20.7	664	12	BG860017 1024066H0
23	94.5	20.7	771	14	CB631692 OS1EB09G
24	94.5	20.7	809	14	CB671698 OSJNE050G
25	94.5	20.7	825	14	CB675995 OSJNE050G
26	94.5	20.7	839	14	CB656383 OSJNE050G
27	94.5	20.7	1343	12	BG284023 602407325
28	94	20.6	282	10	BF364571 PMI-NN108
29	94	20.6	298	29	CE810988 tigr-gss-
30	94	20.6	876	12	BG43231 602541917
31	94	20.6	1114	28	BZ562969 pas62_164
32	93.5	20.5	505	10	B3326969 hr8C04.X
33	93.5	20.5	566	9	AI741157 wg19f07.x
34	93.5	20.5	579	13	BQ786882 il51a04.x
35	93.5	20.5	594	28	BH782442 f2mb011f0
36	93.5	20.5	681	29	AY406899 Homo sapi
37	93.5	20.5	686	10	AM968685 EST380761
38	93	20.4	349	29	CE374135 tigr-gss-
39	93	20.4	545	12	B1849019 471654 MA
40	93	20.4	585	9	AI405829 GH25967.5
41	93	20.4	845	10	BF696818 602125353
42	93	20.4	849	29	CNS0329D Tetraodon
43	93	20.4	872	13	BX325065 BX325065
44	92.5	20.3	465	12	B1255478 602977989
45	92.5	20.3	651	28	CC419125 FUHHS48TB

ALIGNMENTS

RESULT 1  
BC039500  
LOCUS  
DEFINITION Homo sapiens, clone IMAGE:555626, mRNA.  
ACCESSION BC039500  
VERSION BC039500.1 GI:25058499  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1467)  
AUTHORS Strausberg,R.

**TITLE**  
JOURNAL  
Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK**  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdepaxil.stanford.edu](mailto:mcdepaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 88 Row: g Column: 1  
This clone has the following problem: retained intron.

#### FEATURES

Location/Qualifiers  
1. 1467  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:555626"  
/tissue\_type="Uterus, leiomyosarcoma"  
/clone\_lib="NIH MGC 71"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

#### ORIGIN

Alignment Scores:  
Pred. No.: 3.59e-25 Length: 1467  
Score: 456.00 Matches: 79  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 11 Gaps: 0

US-09-972-032-2 (1-79) x BC039500 (1-1467)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisThr 20  
Db 673 ATGTGGGAGACCGCTCGCTAAGCGTGGATGGCTTCGCTGATGCACATTGGACC 732

QY 21 GlyLeuThrGlyLeuGlyGlnGlyGlyLeuGlyGlyLeuGlyGlnAla 40  
Db 733 GGCTCTGACATGGCTAGGGAAGGAGGAGGCGGAATTGGCGCCGAGGCGCC 792

QY 41 SerProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPProThrSer 60  
Db 793 TCGCCGACCCCGACTCGCCCTCCCGTGGCCCGGAGCGCCCTCCCGTGGCCCTCGAGT 852

QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79  
Db 853 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGCGTGGC 909

RESULT 2  
BX406138/c  
LOCUS  
DEFINITION  
BX406138 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
CSODM010YD19 3-PRIME, mRNA sequence.  
ACCESSION  
BX406138  
VERSION  
BX406138.1 GI:30648317  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)

#### AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 8395.f For  
more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOAM010CB1ONP1&cluster=8395.f>. Contact :  
Feng Liang Email: [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSOAM010CB1ONP1.

#### FEATURES

Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODM010YD19"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/note="Organ: liver; Vector: pCMVSPORT6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

#### ORIGIN

Alignment Scores:  
Pred. No.: 1.38e-22 Length: 1201  
Score: 421.00 Matches: 77  
Percent Similarity: 97.47% Conservative: 0  
Best Local Similarity: 97.47% Mismatches: 2  
Query Match: 92.32% Indels: 1  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BX406138 (1-1201)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisThr 20  
Db 713 ATGTGTGGGAGACCGCTCGCTAAGCGTGGATGGTTCGCTGATGCACATTGGACC 654

QY 21 GlyLeuThrThrGlyLeuGlyGlnGlyGlnGlyGlyLeuGlyGlnAla 40  
Db 653 GGGCTCTGGACTGGGCTAGGGAAGGAGGAGGCGGAATTGGCCCGAG-GGCCAGGCC 595

QY 41 SerProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPProThrSer 60  
Db 594 TCGCCGACCCCGACTCGCCCTCCCGTGGCCCGGAGCGCCCTCCCGTGGCCCTCGAGT 535

QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79  
Db 534 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGCGTGGC 478

RESULT 3  
BM980506/c  
LOCUS  
DEFINITION  
BM980506 add-a-08-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
UI-CF-EN1 add-a-08-0-UI 3', mRNA sequence.  
ACCESSION  
BM980506  
VERSION  
BM980506.1 GI:19602038  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 760)  
AUTHORS  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE  
Normalization and subtraction: two approaches to facilitate gene  
discovery

```
ORIGIN
Alignment Scores:
Pred. No.:      1,45e-11          Length:      691
Score:          274.00           Matches:     52
Percent Similarity: 98.11%       Conservatives: 0
Best Local Similarity: 98.11%    Mismatches:   1
Query Match:     60.09%         Indels:      1
DB:              13            Gaps:        0

US-09-972-032-2 (1-79) x BU620427 (1-691)

Qy      27 GlyGluGlyClnGlcGlyIleGlyProGluGlyGlnAlaSerProThrProAspCys 46
|||||
```

```

Db      689  GGGAGAGGCGAGGAGCGGGAAT-GGGCCCCAGAGCGCGCTCGCCGACCCCGACTGC 631
Qy      47  AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
Db      630  GCCTCCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGAGTGCAGGTCTTACCGTCCGA 571
Qy      67  AspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db      570  GATCGTCGCACTGGCGGAGCTGTGCATGGCGGTGC 532

RESULT 5
BU684594/c
LOCUS      BU684594
DEFINITION UI-CF-EN1-act-a-22-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
VERSION     BU684594
KEYWORDS    BU684594.1 GI:23537704
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES             Location/Qualifiers
     source            1..694
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="UI-CF-EN1-act-a-22-0-UI"
     /tissue_type="Primary Lung Cystic Fibrosis Epithelial
     Cells"
     /dev_stage="Adult"
     /lab_host="DH10B (Life Technologies) (71 phage resistant)"
     /clone_lib="UI-CF-EN1"
     /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
     modified polylinker; Site 1: EcoR I; Site 2: Not I;
     UI-CF-EN1 is a normalized cDNA library containing the
     following tissue(s): Primary Lung Cystic Fibrosis
     Epithelial Cells. The library was constructed according to
     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
     1996. First strand cDNA synthesis was primed with an
     oligo-dT primer containing a Not I site. Double stranded
     cDNA was ligated to an EcoR I adaptor, digested with Not
     I, and cloned directionally into pT7T3-Pac vector. The
     oligonucleotide used to prime the synthesis of
     first-strand cDNA contains a library tag sequence that is
     located between the Not I site and the (dT)18 tail. The
     sequence tag for this library is CTGCTCAGGT.
     TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
     6hr to LPS 24h
     TAG_LIB=UI-CF-EN1

```

## TAG\_SEQ=CTGCTCAGGT"

## ORIGIN

```

Alignment Scores:
Pred. No.:      7,248-10      Length:      694
Score:          252.00        Matches:    48
Percent Similarity: 96.00%    Conservative: 0
Best Local Similarity: 96.00% Mismatches:    2
Query Match:    55.26%      Indels:     1
DB:             13          Gaps:      0

```

US-09-972-032-2 (1-79) x BU684594 (1-694)

```

Qy      30  GlnGluGlyGlyTleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArg 49
Db      694  CAGGAGGGCGGAAT-GGGCCCCAGAGCGCGCTCGCCGACCCCGANTGGCGCTCCCGG 636
Qy      50  TrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgPro 69
Db      635  TGGCCCCCGAGCGCTCCCGTGGCCCTGGAGTGCAGTCTTACCGTCCGAGATCGTCCG 576
Qy      70  GlnLeuGlyGluLeuCysMetGlyArgGly 79
Db      575  CAACTGGCGGAGCTGTGCATGGGCGGTGC 546

```

## RESULT 6

BM806108

LOCUS

DEFINITION

AGENCOURT 6553891 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5555626

5', mRNA sequence.

ACCESSION

BM806108

VERSION

BM806108.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1084)

AUTHORS

NIH-MGC http://mgi.nci.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12276 row: b column: 11

High quality sequence stop: 672.

Location/Qualifiers

1..1084

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5555626"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_71"

/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 2.1 kb."

Average insert size 2.1 kb."

## ORIGIN

Alignment Scores:

Pred. No.: 2,048-09 Length: 1084

Score: 249.50 Matches: 61

Percent Similarity: 72.41% Conservative: 2

Best Local Similarity: 70.11% Mismatches: 16

Query Match: 54.71% Indels: 8

DB: 12 Gaps: 3



US-09-972-032-2 (1-79) x BM806108 (1-1084)

Qy 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaaspAlaHisTrpThr 20  
 Db 672 ATGTGTGGGAGACCGGTCCGTAGCCCTGGATGGCTTCCTGATGACATTGGACC 731

Qy 21 GlyLeuTrpThrGlyLeuGlyGlu-GlyGlnGluGlyGlyLe----GlyProGluGlyG 39  
 Db 732 GGGCTCGGACTGGGCGCTAGGGGAAGGCCAGAGGGCCGATTTGGGGGCGCCGAGGGCCA 791

Qy 39 nAlaSerProThrProAspCysAla-SerArgTrp-ProArgSerAlaSerArgTrp--P 58  
 Db 792 GGCCTCCGACACCCCGCCCTCCGCGTGGGCGCCGACAGGGCGCTCCCGTGGGCC 851

Qy 58 roTrpSerAla---GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCys---M 76  
 Db 852 CCTGGGAGGCAAGGCTTTACCGTCCCAATCGTCCCAACCTGGCGCAATTGGGCT 911

Qy 76 etGlyArgGly 79  
 Db 912 GGGGGCGGGGC 922

RESULT 7  
 BM679577/c  
 LOCUS BM679577 669 bp mRNA linear EST 27-FEB-2002  
 DEFINITION UI-E-E00-aia-1-05-0-UI.s1 UI-E-E00 Homo sapiens cDNA clone  
 UI-E-E00-aia-1-05-0-UI 3', mRNA sequence.

ACCESSION BM679577  
 VERSION BM679577.1 GI:18989473  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 669)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bentos@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES  
 source Location/Qualifiers  
 1..669  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-E00-aia-1-05-0-UI"  
 /tissue\_type="fetal eye"  
 /dev stage="fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-E00"

with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p77T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (G)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG\_TISSUE=human fetal eye  
 TAG\_LIB=UI-E-E00  
 TAG\_SEQ=CGGTATACC"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 8,29e-09 Length: 669  
 Score: 238.00 Matches: 42  
 Percent Similarity: 97.6% Conservative: 0  
 Best Local Similarity: 97.6% Mismatches: 1  
 Query Match: 52.1% Indels: 0  
 DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM679577 (1-669)

Qy 37 GluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56  
 Db 669 GAGGCGCCAGGCTCCGCCACCCCGANTGCCCTCCGCGCCGCGCCGCGCCGCG 610

Qy 57 TrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMet 76  
 Db 609 TGGCCCTGGAGTGCAGGCTTACCGTCCGAGATCGTCCGCACTGGCGAGCTGTGCATG 550

Qy 77 GlyArgGly 79  
 Db 549 GGGCGTGGC 541

RESULT 8  
 BU617331/c  
 LOCUS BU617331 632 bp mRNA linear EST 23-SEP-2002  
 DEFINITION UI-H-DF0-bep-n-09-0-UI.s1 NCI-CGAP\_DFO Homo sapiens cDNA clone  
 UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.

ACCESSION BU617331  
 VERSION BU617331.1 GI:23283539  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 632)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@mail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bentos@uiowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source Location/Qualifiers  
 1..632  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-DF0-bep-n-09-0-UI"  
 /tissue\_type="Subchondral Bone"  
 /dev stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI-CGAP\_DFO"

/note="Organ: Bone; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP DFO is a cDNA library containing the following tissue(s): Subchondral Bone. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.  
TAG\_TISSUE=subchondral bone  
TAG\_LIB=UI-H-DFO  
TAG\_SEQ=GTTAAGCGTC"

## ORIGIN

Alignment Scores:  
Pred. No.: 6.6e-05 Length: 632  
Score: 187.00 Matches: 33  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 41.01% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BU617331 (1-632)

Qy 47 AlaserArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66  
Db 630 GCCTCCCGGTGGCCCGCAGCGCTCCCGTGGAGTGCAGGTCTTACCGTCCGA 571  
Qy 67 AsparGProGlnLeuGlyGluLeuCysMetGlyArgGly 79  
Db 570 GATCGTCCGCACTGGCGGAGCTGTGCATGGGCGGTGGC 532

## RESULT 9

CA313142/c  
LOCUS CA313142 633 bp mRNA linear EST 04-NOV-2002  
DEFINITION UI-CF-FNO-aex-n-23-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone  
UI-CF-FNO-aex-n-23-0-UI 3', mRNA sequence.

ACCESSION CA313142.1 GI:24531240  
VERSION CA313142.1  
KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 633)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 889548

COMMENT Contact: McCray, PB

McCrack Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com) or from Open Biosystems

(www.openbiosystems.com).

Seq primer: M13 FORWARD

POLYA=Yes.

## FEATURES

source

1. .633

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-CF-FNO-aex-n-23-0-UI"

/tissue\_type="Human Lung Epithelial cells"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-CF-FNO"

/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-FNO is a subtracted cDNA library derived from two

normalized Human lung epithelial cell libraries (EN1 and

DUI) The library was subtracted according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. For additional information, contact:

bento-soares@uiowa.edu

TAG\_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG\_LIB=UI-CF-FNO

TAG\_SEQ=CTGCTCAGGT"

## ORIGIN

Alignment Scores:  
Pred. No.: 6.61e-05 Length: 633  
Score: 187.00 Matches: 33  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 41.01% Indels: 0  
DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CA313142 (1-633)

Qy 47 AlaserArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66  
Db 633 GCCTCCCGGTGGCCCGCAGCGCTCCCGTGGAGTGCAGGTCTTACCGTCCGA 574  
Qy 67 AsparGProGlnLeuGlyGluLeuCysMetGlyArgGly 79  
Db 573 GATCGTCCGCACTGGCGGAGCTGTGCATGGGCGGTGGC 535

## RESULT 10

CD742581/c

LOCUS CD742581

DEFINITION UI-H-FT2-bj1-i-10-0-UI.s1 NCI CGAP\_F22 Homo sapiens

UI-H-FT2-bj1-i-10-0-UI 3', mRNA sequence.

ACCESSION CD742581.1 GI:32293431

VERSION CD742581

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 623)

AUTHORS NCI-CGAP

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb-remail.nih.gov

Tissue Procurement: Dr. Gary W. Hunninghake, U of I

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/cgap.html

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

1. .623

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-H-FT2-bj1-i-10-0-UI"

/tissue\_type="Aveolar Macrophage"

```

/dev stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_F72"
/notes="Organ: Lung; Vector: p7773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI_CGAP_F72 is a subcloned cDNA library constructed from
a pool of 81 RNA samples from Alveolar Macrophages
challenged with different treatments. The library was
subtracted according to Bernaldo, Lennon and Soares, Genome
Research, 6:791-806, 1996. The tissue was provided by Dr.
Gary W. Hunninghake of the University of Iowa.
TAG_TISSUE=Human Lung Alveolar Macrophage
TAG_LIB=UI-H-FT2
TAG_SEQ=GGCCATGCGC"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 0.000132 Length: 623
Score: 183.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.13% Indels: 0
DB: 14 Gaps: 0

```

US-09-972-032-2 (1-79) x CD7423581 (1-623)

```

Qy 48 SerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAsp 67
|||||
Db 622 TCCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGTCTTACCGTCCGAGAT 563
|||||
Qy 68 ArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
|||||
Db 562 CGTCCGCAACTGGCGAGCTGTGCATGGGCGGTGGC 527
|||||

```

## RESULT 11

```

BQ447041/c
LOCUS
DEFINITION
UT-H-EUI-bac-p-06-0-UI.s1 NCI_CGAP_Ctl Homo sapiens cDNA clone
UT-H-EUI-bac-p-06-0-UI 3', mRNA sequence.
ACCESSION
BQ447041
VERSION
BQ447041.1 GI:21250153
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 628)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..628
/organism="Homo sapiens"
/mol_type="mRNA"
/clone="UI-H-EUI-bac-p-06-0-UI"
/tissue type="Osteoarthritic Cartilage"
/dev stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Ctl"
/notes="Organ: Knee; Vector: p7773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;

```

## FEATURES

```

source
1..628
Location/Qualifiers
1..486
/organism="Sus scrofa"
/mol_type="mRNA"
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 2 row: X column: 16
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..486
/organism="Sus scrofa"
/mol_type="mRNA"

```

NCI\_CGAP\_Ctl is a normalized cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCT.

TAG\_TISSUE=osteoarthritic cartilage  
TAG\_LIB=UI-H-EUI  
TAG\_SEQ=TGATCAGCT"

## ORIGIN

```

Alignment Scores:
Pred. No.: 0.00273 Length: 628
Score: 166.00 Matches: 31
Percent Similarity: 96.88% Conservative: 0
Best Local Similarity: 96.88% Mismatches: 0
Query Match: 36.40% Indels: 1
DB: 13 Gaps: 0

```

US-09-972-032-2 (1-79) x BQ447041 (1-628)

```

Qy 49 ArgTrpProArgSerAlaSerArg-TrpProTrpSerAlaGlyLeuThrValArgAsp 68
|||||
Db 626 CGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGTCTTACCGTCCGAGATCG 567
|||||
Qy 68 gProGlnLeuGlyGluLeuCysMetGlyArgGly 79
|||||
Db 566 TCCGCAACTGGCGAGCTGTGCATGGGCGGTGGC 533
|||||

```

## RESULT 12

```

BQ484011/c
LOCUS
DEFINITION
537701 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION
BQ484011
VERSION
BQ484011.1 GI:18534339
KEYWORDS
EST.
SOURCE
Sus scrofa (pig)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 486)
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Vallet, J., Wise, T., Rohrer, G.A., Pettea, G., Sultana, R.,
Quackenbush, J. and Keele, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 2 row: X column: 16
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..486
/organism="Sus scrofa"
/mol_type="mRNA"

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT



http://image.llnl.gov  
Plate: LLCM1513 row: e column: 14  
High quality sequence stop: 591.  
Location/Qualifiers

## FEATURES

source

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone=IMAGE:4692469  
/lab_host="DHIOB (T1 phage-resistant)"  
/clone_name="NIH MGC 77"  
/note="Organ; Link; Vector: pDNR-LIB (+/-SfiI) (gccgcctccggc); Site 2: SfiI (ggttcacagcgatgcatg-3' overhang) as for SfiI adaptors were used in cloning as follows: sequence: 5'-CAGCGCATATGCCC-3' and 5'-ATTCTAGACGGCAGCGCCGATG-dt(30)BspEI. C, or G and N = A, C, G, or T). Averaging 0.5-4.0 kb (range 0.5-4.0 kb), 12/15 colonies by PCR. This library was enriched for BAC clones and was constructed by Clontech Laboratory (CA). Note: this is a NIH MGC Library."
```

## ORIGIN

Alignment Scores:	
Pred. No.:	1.15e+03
Score:	95.50
Percent Similarity:	36.79%
Best Local Similarity:	30.19%
Query Match:	20.94%
DB:	12
Gaps:	5
Indels:	37
Mismatches:	30
Conservative:	7
Matches:	32
Length:	872

US-09-972-032-2 (1-79) X BG539339 (1-872)

2	CysGlyArgProArgArgValserAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly	21
QY		
207	TGTCCAGACGCTGGAGATGGCGCGAGCGCAGCGTGGGGGTGTGGAGTCTCTGCTGC	266
DB		
22	LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyProGlu	37
QY		
267	TGGTGGACAGGACCCGGGGAGGCCGA-----GCCCTGGGGCCAGAGCCGCCAGGTGG	320
DB		
38	-----GlyIlnAlaSerProThrPro	44
QY		
321	ACGCACATGACTGTGAGGCCCTGGGCTGGGGCTCCGAGGCAGCCGCTCCAGGAGTCC	380
DB		
45	Asp-----CysAlaSerArg-----TrpProArg	52
QY		
381	GAGCACCGTCAAACTGCAGTTCGACATGATCGCGCTTGCACACTGTGTGGCAGCGCG	440
DB		

BG539339

441 CGCTACCGAGGCGACCTTCTCTGGGGCTGTGGCCCTGCCCTGCTGTAC 500

ACCESSION

## KEYWORDS

ORGANISM

**RESEARCH**

[illegible]

COMMENT	DOCTAD



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 10, 2004, 15:31:05 ; Search time 2821 Seconds  
(without alignments)  
1213.789 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 456  
Sequence: 1 MCGRRRVSGAGCFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2 1/USPTO.spool/US09972032/runat\_06072004\_121451\_8389/app query.fasta\_1.263  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972032 @C@N 1 1 3731 @runat\_06072004\_121451\_8389 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.em.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rdi.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	182230	9	AC135050 Homo sapi
2	456	100.0	195476	2	AC135044 Homo sapi
C 3	105	23.0	3628	8	AK110169 Oryza sat
C 4	103	22.6	163035	2	AC006405 Homo sapi
5	100	21.9	900	9	SS3698 tal-1-exons
6	99.5	21.8	136098	9	AC006970 Homo sapi
C 7	98.5	21.6	2163	5	AF164165 Certhidea
C 8	98.5	21.6	153675	2	AC092012 Felis cat
C 9	98	21.5	185997	2	AC134385 Papio anu
10	98	21.5	191292	2	AC138018 Homo sapi
C 11	98	21.5	303550	1	SC0939131 Streptomy
12	97.5	21.4	1476	6	BD179739 Highly th
13	97.5	21.4	45858	9	AC006123 Homo sapi
14	97.5	21.4	283100	1	SC0939110 Streptomy
C 15	97	21.3	193012	2	AC113007 Mus muscu
16	97	21.3	300800	1	SC0939112 Streptomy
17	96	21.1	161090	9	AC117415 Homo sapi
18	96	21.1	193265	2	AC138705 Homo sapi
19	96	21.1	308147	1	AE016915 Chromobac
C 20	95.5	20.9	1325	8	AK106138 Oryza sat
21	95.5	20.9	72907	2	AC016010 Homo sapi
22	95.5	20.9	143961	8	AC018929 Oryza sat
23	95.5	20.9	300029	8	AE017122 Oryza sat
24	95	20.8	44491	9	AC073308 Homo sapi
25	95	20.8	88733	2	AC132855 Homo sapi
26	95	20.8	151495	2	AC068240 Homo sapi
C 27	95	20.8	181671	9	AC007040 Homo sapi
C 28	95	20.8	345997	1	BX569692 Synchoco
C 29	95	20.8	349901	6	AX647201 Sequence
C 30	94.5	20.7	809	4	BTGST
C 31	94.5	20.7	1722	8	AK060525 Oryza sat
32	94.5	20.7	24081	6	AX598605 Sequence
33	94.5	20.7	52101	6	AX598593 Sequence
34	94.5	20.7	131424	8	AF004815 Oryza sat
C 35	94.5	20.7	140241	9	AC079355 Homo sapi
C 36	94.5	20.7	188121	2	AC093394 Bos tauru
C 37	94.5	20.7	227595	2	AC093396 Bos tauru
38	94.5	20.7	230138	14	AF232689 Rat cytom
39	93.5	20.5	606	6	AX119213 Sequence
40	93.5	20.5	943	6	AX574584 Sequence
41	93.5	20.5	943	9	AY358178 Homo sapi
42	93.5	20.5	1641	9	BC004932 Homo sapi
43	93.5	20.5	1870	9	AK024551 Homo sapi
44	93.5	20.5	3161	9	AK126094 Homo sapi
45	93	20.4	45886	2	AC017393 Drosophil

## ALIGNMENTS

RESULT 1

```

AC135050      182230 bp      DNA      linear      PRI 27-FEB-2003
LOCUS      Homo sapiens chromosome 16 clone RP11-196G11, complete sequence.
ACCESSION      AC135050
VERSION      AC135050.3 GI:28570306
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
REFERENCE      1 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      3 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (22-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      4 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Submitted (27-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT      On Feb 27, 2003 this sequence version replaced gi:24211094.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
FEATURES      Location/Qualifiers
             1..182230
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="16"
                /clone="RP11-196G11"
ORIGIN
Alignment Scores:
Pred. No.:      5,23e-21      Length:      182230
Score:      456.00      Matches:      79
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      9      Gaps:      0
US-09-972-032-2 (1-79) x AC135050 (1-182230)
Qy      1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
      |||
Db      98089 ATGTGTGGAGACCGCGTCGTAAGCGTGGATGTGGCTTCGCTCATGCACATTGCACC 98148
      |||
Qy      21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlnAla 40
      |||
Db      98149 GGGCTCTGGAGTGGCTAGGGGAAGGGCAGAGGGGGGGAATTGGGGCCCGAGGCCAGGCC 98208
      |||
Qy      41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSer 60
      |||
Db      98209 TCGCCGACCCCGGACTCGGCTCCCGGTGGCCCCCGCAGCGCCCTCCCGGTGGCCCTGAGT 98268
      |||
Qy      61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79
      |||
Db      98269 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGGTGGC 98325

```

```

RESULT 2
AC135044
LOCUS
DEFINITION
ACCESSION
AC135044
VERSION
AC135044.1 GI:23505535
KEYWORDS
HTG; HTGS_PHASBI; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute.
REFERENCE      1 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center Clone Name: C1PB-EL_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Estimated insert size: 193376; sum-of-contigs estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1      1202: contig of 1202 bp in length
*      1203      1302: gap of unknown length
*      1303      2468: contig of 1166 bp in length
*      2469      2568: gap of unknown length
*      2569      4077: contig of 1509 bp in length
*      4078      4177: gap of unknown length
*      4178      5772: contig of 1595 bp in length
*      5773      5872: gap of unknown length
*      5873      7722: contig of 1850 bp in length
*      7723      7822: gap of unknown length
*      7823      8958: contig of 1136 bp in length
*      8959      9058: gap of unknown length
*      9059      11051: contig of 1993 bp in length
*      11052      11151: gap of unknown length
*      11152      13641: contig of 2490 bp in length
*      13642      13741: gap of unknown length
*      13743      17738: contig of 3997 bp in length
*      17739      17838: gap of unknown length
*      17839      22030: contig of 4192 bp in length
*      22031      22130: gap of unknown length
*      22131      27002: contig of 4872 bp in length
*      27003      27102: gap of unknown length
*      27103      32819: contig of 5717 bp in length
*      32820      32919: gap of unknown length
*      32920      39003: contig of 6084 bp in length
*      39004      39103: gap of unknown length

```





```

source      1. 3628
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="002-161-G03"

ORIGIN
Alignment Scores:
Pred. No.: 170          Length: 3628
Score: 105.00          Matches: 38
Percent Similarity: 39.09%      Conservative: 5
Best Local Similarity: 34.55%    Mismatches: 25
Query Match: 23.03%             Indels: 42
DB: 8                          Gaps: 9

US-09-972-032-2 (1-79) x AK110169 (1-3628)
Qy 2 CysGlyArg-----ProArgArgValSerAla-----10
Db 998 TCGGTCGTACTGCATCGCGGTCTGGACGACCTCGGCACCTCTGTCGGCGCTTTGTAGG 939
Qy 11 GlyCysGlyPheAlaAspAlaH:stTrrThrGlyLeuTrrThrGlyLeuGly-----27
Db 938 GGACCTGGA-----TGGACTGTGTAGTTGGACGATCGAGGTTGGTGACCA 894
Qy 28 GluGly---GlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCys 46
Db 893 GAAGTTGGCGAGGAGCGCTGGAGCTGCTGGACGACATATGAGCGCGCTTGTCT 834
Qy 47 AlaSerArgTrrProArgSerAlaSerArgTrrProTrr-----59
Db 833 CGAGTTCGTTGG---CGATGACGAGACCGTGGCATGGGTTGTAGTGTCTGGGCGCG 777
Qy 60 -----SerAlaGlyLeuThr-----ValArgAspArgPro 69
Db 776 AGTGAAGGCGAGTGGAGCGATCGATGAGTGTGCGGTCTTGGATCCGGAGCGG----720
Qy 70 GlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 719 -----CGCACATGTCTGAGCAGCAGGT 699

RESULT 4
AC006405/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone hRPK.107 N.19 map 17, ***
SEQUENCING IN PROGRESS ***, 12 unordered pieces.
AC006405
AC006405.1 GI:4165361
HTG; HTGS_PHASE1.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 163035)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Baria,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,

```

```

Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 21, 1999 this sequence version replaced gi:4159875.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8815: contig of 8815 bp in length
* gap of unknown length
* 8816 12285: contig of 3470 bp in length
* gap of unknown length
* 12286 40956: contig of 28671 bp in length
* gap of unknown length
* 40957 64393: contig of 23437 bp in length
* gap of unknown length
* 64394 78975: contig of 14582 bp in length
* gap of unknown length
* 78976 96637: contig of 17662 bp in length
* gap of unknown length
* 96638 105967: contig of 9330 bp in length
* gap of unknown length
* 105968 145832: contig of 39865 bp in length
* gap of unknown length
* 145833 147041: contig of 1209 bp in length
* gap of unknown length
* 147042 150498: contig of 3457 bp in length
* gap of unknown length
* 150499 160979: contig of 10481 bp in length
* gap of unknown length
* 160980 163035: contig of 2056 bp in length.
FEATURES
Location/Qualifiers
1..163035
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.107 N.19"
/clone_lib="RPC1-11 human BAC library"
ORIGIN
Alignment Scores:
Pred. No.: 5.39e+03          Length: 163035
Score: 103.00          Matches: 38
Percent Similarity: 37.74%      Conservative: 2
Best Local Similarity: 35.85%    Mismatches: 33
Query Match: 22.59%             Indels: 34
DB: 2                          Gaps: 5

US-09-972-032-2 (1-79) x AC006405 (1-163035)
Qy 3 GlyArgProArgArgValSerAlaGlyCysGlyPhe-----AlaAspAlaHisTrr 19
Db 145698 GGTTCGGCCGGGGGTGTAGGGGAGCGGGGGCGTTCGCATCTGAGCGGGGCCAC----145642
Qy 20 ThrGlyLeuTrrThrGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 39
Db 145641 ---GGGCGATGACGCGGC---GGCCGGGGCATGAGGGGGGCCACGGGAGGACAA 145588
Qy 40 AlaSerProThrProAspCysAlaSerArgTrrProArgSerAlaSerArgTrrProTrr 59
Db 145587 CCGGGCGGACGGCTCTCTAGCGGCCCGGGGGCCCTCGG---AGGCCAAGTGGAGCCG-145532
Qy 60 SerAlaGlyLeuThrValArgAspArgProGln-----70

```

Db 145531 AGGAGGCGGGGACAGCGAGGCGGAGCCAGCGGCTCAGGCGAGGAGA 145472  
Qy 71 -----LeuGlyGlu 73  
Db 145471 GCTGCGCGGCGACTGGGGGAGCGGAGCGGCGGAGCGCGCTGGAGTGTGTGGGTCN 145412  
Qy 74 LeuCysMetGlyArgely 79  
Db 145411 CTCTGCGCGGCGGAGG 145394  
RESULT 5  
S53698  
LOCUS  
DEFINITION  
tal-1-exons 1a, 1b and flanking DNA [human, HEL cell line, Genomic,  
900 nt].  
ACCESSION  
S53698  
VERSION  
S53698.1 GI:234760  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 900)  
AUTHORS  
Bernard,O., Lecointe,N., Jonveaux,P., Souvri,M., Mauchauffe,M.,  
Berger,R., Larsen,C.J. and Mathieu-Mahul,D.  
TITLE  
Two site-specific deletions and t(1;14) translocation restricted to  
human T-cell acute leukemias disrupt the 5' part of the tal-1 gene  
JOURNAL  
Oncogene 6 (8), 1477-1488 (1991)  
MEDLINE  
91360285  
PUBMED  
1886719  
REMARK  
GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 53698] from the original journal article.  
COMMENT  
Exon 1a (nt 269-379) and exon 1b (nt 657-890) are involved in  
alternative splicing of tal-1 gene.  
FEATURES  
source  
1..900  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
gene  
1..900  
/gene="tal-1"  
/note="exons 1a, 1b and flanking DNA"  
ORIGIN  
Alignment Scores:  
Pred No.: 118 Length: 900  
Score: 100.00 Matches: 28  
Percent Similarity: 41.25% Conservative: 5  
Best Local Similarity: 35.00% Mismatches: 31  
Query Match: 21.93% Indels: 16  
DB: 9 Gaps: 4  
US-09-972-032-2 (1-79) x S53698 (1-900)  
Qy 4 ArgProArgArgValSerAlaGlyCys-----GlyPheAlaAsp 16  
Db 195 CGGCGCGCGCGCGCTCGCGCCATGCTAACGAGCCCTGGAACCTGTGAGTGGGATTACAGCG 254  
Qy 17 AlaHisTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGlyGlyPro 36  
Db 255 CGTGGTGTAGGAGCGCG---CGCGGCGCGGGCGCGGGAGCGGGGCGCGCGG--- 308  
Qy 37 GluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56  
Db 309 AGGGGCGGGCGCGCGCGCTCAGACCGGGCCCTCAAAATGGCCACACGCGGTACCCCGT 368  
Qy 57 -----TTPProTrpSerIaGlyLeuThrValArgAspArgPro 69  
Db 369 AGCGGAAAAACGTGAGCATTTCTGCGCTTTTCTAGGGGAAAAGCAACCGCGGCGCTCCCA 428  
RESULT 6

AC006970  
LOCUS  
DEFINITION  
Homo sapiens PAC clone RP4-725G10 from 7, complete sequence.  
AC006970  
VERSION  
AC006970.6 GI:10312289  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 136098)  
AUTHORS  
Sulston,J.E. and Wilson,R.  
TITLE  
Toward a complete human genome sequence  
JOURNAL  
Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE  
99063792  
PUBMED  
9847074  
REFERENCE  
2 (bases 1 to 136098)  
AUTHORS  
Du,H., Haakenson,B. and Stoneking,T.  
TITLE  
The sequence of Homo sapiens PAC clone RP4-725G10  
JOURNAL  
Unpublished (2001)  
REFERENCE  
3 (bases 1 to 136098)  
AUTHORS  
Waterston,R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (05-MAR-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE  
4 (bases 1 to 136098)  
AUTHORS  
Waterston,R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (27-SEP-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE  
5 (bases 1 to 136098)  
AUTHORS  
Wilson,R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (15-OCT-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT  
On Sep 27, 2000 this sequence version replaced gi:9838025.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: sapiens@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_DJ0725G10  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu

SOURCE INFORMATION:  
This clone was derived from human PAC library RPI-4, prepared by Pieter de Jong and coworkers at http://www.chori.org using the

method described by Ioannou et al., Nature Genetics 6:84-9 (1994).  
The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-782E10 the clone sequenced  
to the right is RP11-700P18. Actual start of this clone is at base  
position 1 of RP4-725G10 actual end is at base position 136098 of  
RP4-725G10.

## FEATURES

```

source      Location/Qualifiers
1..136098
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP4-725G10"
/clone_lib="RPCI-4"
1..112
/rpt_family="Alu"
113..422
/rpt_family="Alu"
424..716
/rpt_family="Alu"
731..1039
/rpt_family="Alu"
1042..1381
/rpt_family="Alu"
1385..1697
/rpt_family="Alu"
1704..2008
/rpt_family="Alu"
2347..2557
/rpt_family="Alu"
2558..2693
/rpt_family="Alu"
2694..2999
/rpt_family="Alu"
3000..3175
/rpt_family="Alu"
3176..3299
/rpt_family="Alu"
3384..3469
/rpt_family="MIR"
3489..3788
/rpt_family="Alu"
3804..3829
/rpt_family="AT-rich"
4018..4187
/rpt_family="Alu"
4188..4491
/rpt_family="Alu"
4492..4660
/rpt_family="Alu"
4664..4963
/rpt_family="Alu"
5048..5287
/rpt_family="MER2_type"
5299..5476
/rpt_family="Alu"
5480..5545
/rpt_family="MER2_type"
6045..6161
/rpt_family="Alu"
6195..6292
/rpt_family="U13"
6467..6643
/rpt_family="MER1_type"
6756..6911
/rpt_family="Alu"

```

```

gene      complement(7106..12011)
          /gene="LOC51142"
mRNA      complement(join(7106..7398,8404..8548,9763..10012,
          11901..11975))
          /gene="LOC51142"
mRNA      complement(join(7110..7398,8404..8548,9763..10012,
          11901..11987))
          /gene="LOC51142"
mRNA      complement(join(7112..7398,8404..8548,9763..10012,
          11901..12011))
          /gene="LOC51142"
CDS       complement(join(7388..7398,8404..8548,9763..10012,
          11901..11950))
          /gene="LOC51142"
          /note="Homo sapiens 16.7Kd protein (LOC51142), mRNA.;"
          H_DJ0725G10.1
          /codon_start=1
          /product="unknown"
          /protein_id="AAQ96886.1"
          /db_xref="GI:37674436"
          /translation="MPGRSRRTSRMAPPASRAPQRAAPPAVPAOPPAAPPASVAVG
          SSAAPRQPGIMQAVATTAAGVAVGSAVGHILGHTGGSGGSNAAPAPDITYQBP
          QGTQPAQQQPCLYEIKQFLCAQNGQDKLCEGFNEVLKQCRILANGLA"
          7504..7631
          /rpt_family="Alu"
          repeat_region
          7883..8204
          /rpt_family="Alu"
          repeat_region
          8735..8952
          /rpt_family="Alu"
          repeat_region
          9004..9129
          /rpt_family="Alu"
          repeat_region
          10149..10446
          /rpt_family="Alu"
          repeat_region
          10450..10747
          /rpt_family="Alu"
          repeat_region
          10941..11027
          /rpt_family="L2"
          repeat_region
          11030..11337
          /rpt_family="Alu"
          repeat_region
          11372..11440
          /rpt_family="MIR"
          misc_feature
          11557..12242
          /note="CpG island (GC=64.2, o/e=1.07, #CpGs=48)"
          12361..12400
          /rpt_family="MIR"
          repeat_region
          12823..12946
          /rpt_family="L2"
          repeat_region
          12980..13272
          /rpt_family="Alu"
          repeat_region
          13523..13750
          /rpt_family="L1"
          repeat_region
          14051..14141
          /rpt_family="L1"
          repeat_region
          14142..14441
          /rpt_family="Alu"

```

## Alignment Scores:

```

Pred. No.:      8.04e+03      Length:      136098
Score:          99.50         Matches:      33
Percent Similarity: 32.43%    Conservative: 3
Best Local Similarity: 29.73% Mismatches:      39
Query Match:     21.82%       Indels:       36
DB:              9           Gaps:         4

```

US-09-972-032-2 (1-79) x AC006970 (1-136098)

```

Qy      3 GlyArgProArgArg-ValSerAlaGlyCysGlyPheAlaAspAlaHisItrThrGlyLe 22
          |||||
Db      80356 GGAAGTCCAGCGCGGTGCAAGCGGTTCGCGCGTGGCGCGCGCGCGAGTGC 80415
          |||||
Qy      22 uTrpThrGlyLeuGly-----GluGlyGlnGluGlyGlyLeGly----- 35
          |||||
Db      80416 GCGGGTTGGGCTAGGGGCTGTCCTCGAGCGCGGTGGCCCTCGCGCACACCCAC 80475
          |||||

```







```
/mol_type="genomic DNA"
/strain="A3[2]"
/db_xref="taxon:100226"
135..1643
/gene="SC07252"
/notes="synonym: SC5H1.40c"
135..1643
/gene="SC07252"
/notes="SC5H1.40c, possible regulatory protein, len: 502
aa; similar to e.g. SW:SP15_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores; opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPTREMBL:Q9X7U7"
/translation="VSEPNTRLSDLFLAGWSKGLARLVNQAANGHPOLATDTS
RVRWIDMGRIPDEVPRLAALFTEGLGVTVLEDLGLVHCRAGKRPDSEERPD
GVWPAPRTAAVLTEFTGMDLMLNRGLVGAGALAAGSALSAMDWLTDPALAAD
APDHLQHPADPGFDYEAAPIGSQSEVESLEKSVFRAWDAGGGQKRAVGGQL
NEVGGMLAYHPHQLRRLGVAANAVLAGWSDHVGLEPTAKYFVIAAHAAREGG
DRPAGALSRARQMVHLGKDPDALMLKQSGSGEQLVPRTKMLYTIEMAAQAS
MGKQAMRTILGRAEDLVSDKADVPDMMQTFKSEDIYGMQALAYRTLAEFPGAA
AAQCHAYADKALARIQGRORSKIFDYLMSACFIADDPBOADRYARLALVMSGNS
QRTWDLRLQMYLTAEGYPKIQELREELKALPKGSRGKPGGTAPA"
complement(1663..2979)
/gene="SC07253"
/notes="synonym: SC5H1.39"
complement(1663..2979)
/gene="SC07253"
/notes="SC5H1.39, hypothetical protein, len: 438 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame analysis and amino acid
composition"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB42964.1"
/db_xref="GI:4835342"
/db_xref="GOA:Q9X7U6"
/db_xref="SPTREMBL:Q9X7U6"
/translation="VSSHLSDHPAAQSQGQSVREALISQTRLRGVDVARRDSQDD
TTDPGRWQALYDLALHQLTDLDAHLAQLRDGPAPAPRAGSLLSRVGAENWLLTDE
AENGAELPHILGRDPAVPLSLDELPSVLTEDRPKLTAMVTDCLVDGRPIDGEFRV
PDGASVTVMMGCEPVLADGSTAMWAVLRDVELRPSORAVSETRDSLRPPASPW
HGLLRFDDEGRGLDLGAHLPTSATGRSREGGWYDAGPLGADGELITLVGDEGLP
EADYGAVALTGALRGDLAAGTAPRGLRSLAELLKATALAPVDGAVYGYRPDTHLL
VEHLTLAPRLTADSGDGVRLVITRELDGFPGRQACVLIAVRTP"
complement(3252..4280)
/gene="SC07254"
/notes="synonym: SC5H1.38"
complement(3252..4280)
/gene="SC07254"
/notes="SC5H1.38, probable myo-inositol dehydrogenase, len:
342 aa; similar to many e.g. SW:M12D_BACSU (EMBL:M76431),
idh, Bacillus subtilis myo-inositol 2-dehydrogenase (344
aa), fasta scores; opt: 864 z-score: 958.5 E(): 0.38.1%
identity in 336 aa overlap. Also similar to SW:STRI_STRGR
(EMBL:Y00459), stri, Streptomyces griseus streptomycin
biosynthesis myo-inositol oxidoreductase (348 aa) (32.7%
identity in 343 aa overlap). Weakly similar to Tr:O69945
(EMBL:AL023862) S.coelicolor possible oxidoreductase (430
aa) (29.8% identity in 228 aa overlap). Contains Pfam
match to entry PF01408 GFO IDH Moca, Oxidoreductase
family, score 153.50, E-value 3.6e-42"
/codon_start=1
/transl_table=11
/product="putative myo-inositol dehydrogenase"
/protein_id="CAB42963.1"
/db_xref="GI:4835341"
/db_xref="GOA:Q9X7U5"
/db_xref="SPTREMBL:Q9X7U5"
/translation="MSBELLGVALGAGHMGADHIRRVDVVVGARVAADPPDAERAK
EAVGIGGTGRITVHTDVEAALDAPGVEAVLIASPGEAHEALAAAFGLVLCCEKP
MAPNSAGALRVVEAEARLQAQIGFMRRYDAEYRQLKSLDGLGFRPLMLRCHVHR
NVSPPHFTSAMLINSSVSHDEIDARLLIGQELSAVTVLRPRPSAGAPGLLDQLVL
PETGGAVDVEVFNQGFYVCEAVCEAGSARIGAAHTMTVTAAGAAEEVEPODY
LVRRDADYREVQSWDATRGLVTGPTWDGYAAAAGVRAJLDTGVTGRTVDMAP
RPSUHDRA"
complement(3549..4268)
/gene="SC07254"
/notes="Pfam match to entry PF01408 GFO IDH Moca,
Oxidoreductase family, score 153.50, E-value 3.6e-42"
complement(4291..4295)
/notes="possible RBS"
complement(4404..5498)
/gene="SC07255"
/notes="synonym: SC5H1.37"
complement(4404..5498)
/gene="SC07255"
/notes="SC07255"
/notes="SC5H1.37, hypothetical protein, len: 364 aa;
unknown function, similar to Tr:P73120 (EMBL:D90903)
Synchocystis sp. hypothetical protein (358 aa), fasta
scores; opt: 492 z-score: 559.6 E(): 7.6e-24, 31.6%
identity in 364 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB42962.1"
/db_xref="GI:4835340"
/db_xref="SPTREMBL:Q9X7U4"
/translation="MEHDLSPATLGLRRPRPYPYAVSVLTPTTHREPDNAQRVLRN
AVAAKOLEEDPAVSDRAEVSLELDRLAEIDLTVAEDGLAIFAAPGHEQVWTLIS
RSPERPVLSDTFLTNLVAQAARPPFWLSVAADVKLMSGGAGRVHEHLGFFPL
DRPADDFPERQERIGDAPSTFRDEGRRLRDADAMGRILRAHPRLVYTGPEPL
SLFEAGALASDAVTPHGLAGHTAEAWQVAPVLEKARGSVSAEARELDARGR
KDFAGVDLWESAVTGRVLLAVEENFKVTRDDGHELLFADDDGLDAREDIVDEIV
EQCLETGADVRFPDGGALGDVDYDGIAGVLR"
complement(5513..5516)
/notes="possible RBS"
complement(5580..7211)
/gene="SC07256"
/notes="synonym: SC5H1.36"
complement(5580..7211)
/gene="SC07256"
/notes="SC5H1.36, probable protease (putative secreted
protein), len: 543 aa; similar to many e.g. Tr:Q54398
(EMBL:I42758), StpD, Streptomyces lividans protease, (539
aa), fasta scores; opt: 697 z-score: 769.7 E(): 0.35.8%
identity in 545 aa overlap. Similar to other S.coelicolor
proteases e.g. StpD, SCH5.02c (EMBL:AL03563) proteinase
(539 aa) (35.8% identity in 545 aa overlap). Contains
possible N-terminal signal sequence. Contains Pfam match
to entry PF00561 abhydrolase, alpha/beta hydrolase fold,
score 55.00, E-value 1.6e-12"
/codon_start=1
/transl_table=11
/product="putative protease (putative secreted protein)"
/protein_id="CAB42961.1"
/db_xref="GI:4835339"
/db_xref="GOA:Q9X7U3"
/db_xref="SPTREMBL:Q9X7U3"
/translation="MQAKLETRPSVRCANTGALGLAGLPAASGATTAATRTDT
SGSLRFVDQVETACGETEMPKDQCKGVYFUDRSRERGGLTVVALARYATGDS
RGSVLNFGPGGGINELAGKGFHMLTNGVYDVTDFRGVGRDSVSPGCPATLKI
MEATDYGDMTDPGEVRLERLAAAECAKSGSVPLHGTVAADARDVMRRALGDDR
LNYLGFSGYTRLGAVYARFPDKVGRVLDGVTLTPELAEQGLAGAGQGTALLENL
DWCVSDVACPFQGDARDQVRLVSLDSDVPVAFGFFFTGQDMVGATGQALYSR
ELMPSERALLAQLEDGDTGRLGESSGGVTFPVRAAVNAPVRGDTAGLTDERDVP
DNLPAALMAINCADDPRPTAAQVTSILDLRLARYEDVSPVFGVRLTQVLMCYGRPR
```



GTDIRDDVKDLTARVLLVCTRCGPATPYRTWTETADRLGPSAVVLDNRGEGHTGYA  
SSKCVHRKVDVFLYGLPDDGSSCGFESTGDGGG"  
complement(5910. .6809)  
/gene="SC07256"  
/note="Pfam match to entry PF00561 abhydrolase, alpha/beta  
hydrolase fold, score 55.00, E-value 1.6e-12"  
complement(7217. .7220)  
/note="possible RBS"  
7437. .7440  
/note="possible RBS"  
7445. .8158  
/gene="SC07257"  
/note="synonym: SC5H1.35c"  
7445. .8158  
/gene="SC07257"  
/note="SC5H1.35c, possible secreted protein, len: 237 aa;  
unknown function, probable CDS suggested by positional

Alignment Scores:  
Pred. No.: 1.98e+04 Length: 303550  
Score: 98.00 Matches: 26  
Percent Similarity: 40.79% Conservative: 5  
Best Local Similarity: 34.21% Mismatches: 35  
Query Match: 21.49% Indels: 10  
DB: 1 Gaps: 3

US-09-972-032-2 (1-79) x SC0939131 (1-303550)

Qy 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThr 20  
Db 80456 CGGCGAGCGGTACACCGGAGGTGCGGACATGCTCGCACCGGACGGTCTCTGGAGC 80397

Qy 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 40  
Db 80396 ACGGATGAGTCTACTTCGACGTCGCGTCTCCAGCGGTATCCGACGGTTCGAGTCCGGG 80337

Qy 41 SerPro-----ThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56  
Db 80336 TCGCGGAGTCTGCTGCGACGCTCCACCGCGGTCTGCTGCGCGCTCGCGCGCGCGC 80277

Qy 57 TrpProTrp-----SerAlaGlyLeuThrValArgAspArgPro 69  
Db 80276 TGTGTCGACGCGCGCGGAGTGGCGCGCGCGCGGACCGCGCGC 80229

RESULT 12  
BD179739  
LOCUS 1476 bp DNA linear PAT 15-MAY-2003  
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.  
ACCESSION BD179739  
VERSION BD179739.1 GI:30790657  
KEYWORDS JP 2002325574-A/230.  
SOURCE Thermus thermophilus  
ORGANISM Thermus thermophilus  
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
Thermus.  
1 (bases 1 to 1476)  
Kuramitsu,N. and Yokoyama,S.  
Highly thermophilic bacterium-derived protein and gene encoding it  
Patent: JP 2002325574-A 230 12-NOV-2002;  
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
OS Thermus thermophilus  
PN JP 2002325574-A/230  
PD 12-NOV-2002  
PF 23-FEB-2001 JP 2001116171  
PI NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA  
PC C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21, PC  
C12N5/10,  
PC C12N9/88,C12P21/02,/(C12N9/88,C12R1/01),(C12N15/09,C12R1/01),  
PC (C12P21/02,C12R1/01),C12N15/00,C12N15/00,C12N5/00,  
PC C12R1/01)  
CC Highly thermophilic bacterium-derived protein and gene CC  
encoding it  
FH Key Location/Qualifiers

misc\_feature  
complement(5910. .6809)  
/gene="SC07256"  
/note="Pfam match to entry PF00561 abhydrolase, alpha/beta  
hydrolase fold, score 55.00, E-value 1.6e-12"  
complement(7217. .7220)  
/note="possible RBS"  
7437. .7440  
/note="possible RBS"  
7445. .8158  
/gene="SC07257"  
/note="synonym: SC5H1.35c"  
7445. .8158  
/gene="SC07257"  
/note="SC5H1.35c, possible secreted protein, len: 237 aa;  
unknown function, probable CDS suggested by positional

FEATURES  
source  
FT CDS Location/Qualifiers  
1..1476  
/organism="Thermus thermophilus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:274"

ORIGIN  
Alignment Scores:  
Pred. No.: 262 Length: 1476  
Score: 97.50 Matches: 22  
Percent Similarity: 42.31% Conservative: 0  
Best Local Similarity: 42.31% Mismatches: 23  
Query Match: 21.38% Indels: 7  
DB: 6 Gaps: 2

US-09-972-032-2 (1-79) x BD179739 (1-1476)

Qy 19 TrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 38  
Db 518 TGGACCGGCTTCTGGAGGCGCTTTGGACCGGATCGGCGCTGGCGGGCGCGAGGGG 577

Qy 39 GlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpPro 58  
Db 578 GTGGGGTCC-----TCCTTTTGGACGAGGAGGCGCTTCTCTCTGGCCC 619

Qy 59 ---TrpSerAlaGlyLeuThrValArgAspArgPro 69  
Db 620 GGGTGTCTCCAGGCGCTTGCCTTCCCTCCACCCCT 655

RESULT 13  
AC006123  
LOCUS 45858 bp DNA linear PRI 17-DEC-1998  
DEFINITION Homo sapiens chromosome 19, cosmid R28177, complete sequence.  
ACCESSION AC006123  
VERSION AC006123.1 GI:3970939  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 45858)  
Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,  
Burkhardt-Schultz,K., Gordon,L., Dias,J., Kyle,A., Brower,A.,  
Stilwagen,S., Phan,H., Velasco,N., De,L., Regala,W., Terry,A.,  
Garnes,J., Danganan,L., Erler,A., Christensen,M., Georgescu,A.,  
Avila,J., Liu,S., Attix,C., Andreise,T., Frankheim,M.,  
Amico-Reller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,  
Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Sanders,C., Ow,D.,  
Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.  
Sequence analysis of a 1.9 Mb contig in 19p12 between UBA52 and  
D198455  
Unpublished  
2 (bases 1 to 45858)  
Lamerdin,J.E.  
Direct Submission  
Submitted (05-DEC-1998) Joint Genome Institute, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
3 (bases 1 to 45858)  
Lamerdin,J.E.  
Direct Submission  
Submitted (17-DEC-1998) Joint Genome Institute, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
Map and sequence are oriented from p telomere to centromere. Cosmid  
R28177 overlaps cosmid R24069 (AC004400) to the left from bases 1  
to 7,688 of this accession and overlaps cosmid R30064 (AC003107) to  
the right from bases 33,879 to 45,858. Additional map and sequence  
information may be obtained at:  
http://www-bio.lnlnl.gov/bbrp/genome/genome.html.  
Location/Qualifiers  
1..45858  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

FEATURES  
source  
FT CDS Location/Qualifiers  
1..45858  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"



```
repeat_region 32422..32726
/rpt_family="AluY"
repeat_region 32768..33066
/rpt_family="AluY"
repeat_region 33718..33907
/rpt_family="AluJ"
repeat_region complement(37634..37785)
/rpt_family="MER7A"
repeat_region complement(37787..38087)
/rpt_family="AluSx"
repeat_region complement(38089..38281)
/rpt_family="MER7B"
repeat_region 40367..40663

Alignment Scores:
Pred. No.: 4.48e+03 Length: 45858
Score: 97.50 Matches: 32
Percent Similarity: 38.38% Conservative: 6
Best Local Similarity: 32.32% Mismatches: 24
Query Match: 21.38% Indels: 37
DB: 9 Gaps: 6

US-09-972-032-2 (1-79) x AC006123 (1-45858)
Qy 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr----- 20
Db 13801 GGGCGCCT-----GGGAGCTGCAGTTCCTCCGACCTACCGGAGCCCTTCT 13848
Qy 21 ---GlyLeu-----TrpThrGlyLeuGly 27
Db 13849 CCAGGCGCTCAGCTGTTTCACTCTCGCCCGCTTGTTCAGCAGCATGGGATGGTGGGA 13908
Qy 28 GluGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCysAla 47
Db 13909 GATGGGACA-----CTGATGCCAGTCCAGTCCGCCAGGCTCTGTGTAGG 13953
Qy 48 SerArgTrp-----ProArgSerAlaSerArgTrpProTrp 59
Db 13954 CGAGGCTGGAGCTGCTGTGTCATCCAGGAGCTCTCCAGAGCTAGCCGCTGG----- 14007
Qy 60 SerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArg 78
Db 14008 -----CTCCAGTAGGGGTTTACCTTTGAGTGGCTGTGTGTTGCTGGGCCAG 14055

RESULT 14
LOCUS SCO939110 283100 bp DNA linear BCT 11-FEB-2003
Streptomyces coelicolor A3(2) complete genome; segment 7/29.
AL939110 AL035591 AL079332 AL096743 AL096811 AL096839 AL096844
AL096849 AL109848 AL132644 AL445403 AL513407 AL591322 AL645882
AL939110.1 GI:24413753
VERSION Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L., Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H., Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M., Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S., Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S., Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G., Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL Nature 417 (6895), 141-147 (2002)
MEDLINE 21996410
PUBMED 1200953
REFERENCE 2 (bases 1 to 283100)
AUTHORS Bentley,S.D.
```

TITLE JOURNAL COMMENT FEATURES source

Direct Submission  
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
On or before Oct 26 2002 this sequence version replaced gi:20520819, gi:20520879, gi:20520880, gi:20520919, gi:20520821, gi:20520887, gi:20520888, gi:20520889, gi:20520884, gi:20520704, gi:20520915, gi:20520918.  
Location/Qualifiers  
1..283100  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
59..520  
/gene="SCO1698"  
/note="synonym: SCI30A.19"  
59..520  
/gene="SCO1698"  
/note="SCO1698"  
/note="SCO1698"  
undefined function e.g. TR:P96907 (EMBL:292770)  
hypothetical protein from Mycobacterium tuberculosis (151 aa) fasta scores; opt: 639, z-score: 785.9, E(): 0, (60.8% identity in 148 aa overlap) and SW:NODN RHILV modulation protein from Rhizobium leguminosarum (151 aa) fasta scores; opt: 388, z-score: 482.2, E(): 1.6e-19, (44.8% identity in 145 aa overlap)  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAB46796.1"  
/db\_xref="GI:5441772"  
/db\_xref="GOA:Q9S254"  
/db\_xref="SPTREMBL:Q9S254"  
/translation="MAEPRIFFASAEVKAAGVGEQLGYTDLVLEQKRIIDFAEATGDH QWIVHPEAAAGPFGTIAHGYSLLPLFGPQLIRVGVKNGVNGTNGKVPFSP QVIGSLRATAVITGVSDVIGQISVAVFTVEREGDKPKVCVAESVSRYYL"  
complement(511..1158)  
/gene="SCO1699"  
/note="synonym: SCI30A.20c"  
complement(511..1158)  
/gene="SCO1699"  
/note="SCO1699"  
len: 215 aa; similar to many e.g. TR:CAB39717 (EMBL:AL049485) putative regulatory protein from Streptomyces coelicolor (223 aa) fasta scores; opt: 330, z-score: 413.5, E(): 1.1e-15, (33.8% identity in 216 aa overlap) and TR:AAD13556 (EMBL:AF080235) LanK regulator of landomycin biosynthesis in Streptomyces cyanogener S136 (192 aa) fasta scores; opt: 170, z-score: 218.6, E(): 7.6e-05, (31.8% identity in 154 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family. Contains possible helix-turn-helix motif (+4..32 SD) 43-64aa."  
/codon\_start=1  
/transl\_table=11  
/product="putative transcriptional regulator"  
/protein\_id="CAB46797.1"  
/db\_xref="GI:5441773"  
/db\_xref="GOA:Q9S253"  
/db\_xref="SPTREMBL:Q9S253"  
/translation="MSTAAETGTDIEPWEVTPDAARRLLVAAEFAERGVHATTT RDIAGAGMSPALYIHYKTEKLLHRSIGTHRAVALRSLAAQGGSAERADAV SSFVWHAQRRTARVVOYELDALGPAREEILTRQCDAAVRGIDDDGVAAGEFDV PDVKGTTVLAVLCIDVARVFNWNGPTTPEVGALYADLVLRWGAEPSPGTAQR"  
complement(943..1083)  
/gene="SCO1699"  
/note="Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family, score 60.70, E-value 4e-15."  
complement(1219..1533)  
/gene="SCO1700"  
/note="synonym: SCI30A.21c"

misc\_feature

gene

```

CDS
complement(1219..1533)
/genes="SC01700"
/notes="SCI30A.21c, putative membrane protein, len: 104 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB46798.1"
/db_xref="GI:5441774"
/db_xref="SPTREMBL:Q9S252"
/translations="MSDTPVKQNTAIFYQAVASPAVAMAATAIGIYQLOADAVRA
FLAIVLYLTSFTLAKIIRDRQEVQIVSRVDQARLILVEHDFPKPGGTGAG
RP"

gene
complement(1609..2760)
/genes="SC01701"
/notes="synonym: SCI30A.22c"
complement(1609..2760)
/genes="SC01701"
/notes="SCI30A.22c, probable acyl-CoA dehydrogenase, len:
383 aa; similar to many both prokaryote and eukaryote eg.
SW:ACDB BACSU acyl-CoA dehydrogenase from Bacillus
subtilis (379 aa) fasta scores; opt: 109.1 z-score:
1197.8, E(): 0, (43.8% identity in 377 aa overlap) and
SW:ACDS RAT acyl-CoA dehydrogenase from Rattus norvegicus
(Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2,
E(): 0, (44.8% identity in 372 aa overlap). Contains Pfam
match to entry PF00441 Acyl-CoA_dh, Acyl-CoA
dehydrogenase."
/codon_start=1
/transl_table=11
/product="putative acyl-CoA dehydrogenase"
/protein_id="CAB46799.1"
/db_xref="GI:5441775"
/db_xref="GOA:Q9S251"
/db_xref="SPTREMBL:Q9S251"
/translations="MNLSESEOTAVRQLARDVFEREIAPIHVVEWDRABEVDRLVKK
LGSVFLGIDIDQYSGSGDHLAYCLVTEELRGDSVVRGIVSLGVAKTIAAGW
DEQKRWLPLGLTSYVCGELTEPGTSGDAGNLTRAVRDGDYVNGTWMFTNG
WADAVLLFARSDIADCHQVSALFPTDPTGLTRTHNGKLGRLQAGTAEVLVDVR
VPASMLAPGSGFVSAMALAKGRMSVAAGCVGIAQAALDAARVATAGERQFGKTTA
HQVLQBELISDIADVDAAIRLVADLIDRQPFVASSKAKLFASAAVRAANNA
LOVFGAYGIDEPAGKLLRDARVMTLYEGTSIQIKLVICRALTGVSAP"
complement(1636..2748)
/genes="SC01701"
/notes="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA
dehydrogenase, score 565.20, E-value 4.3e-166."
2850..3491
/genes="SC01702"
/notes="synonym: SCI30A.23"
2850..3491
/genes="SC01702"
/notes="SCI30A.23, possible transcriptional regulator, len:
213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683)
regulatory protein from Streptomyces coelicolor (259 aa)
fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08,
(32.3% identity in 217 aa overlap). Contains Pfam match to
entry PF00440 tetr, Bacterial regulatory proteins, tetr
family and Prosite match to PS01081 Bacterial regulatory
proteins, tetr family signature. Contains possible
helix-turn-helix motif (+5.17 SD) 31-52aa."
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB46800.1"
/db_xref="GI:5441776"
/db_xref="GOA:Q9S250"
/db_xref="SPTREMBL:Q9S250"
/translations="MARPRKLLSTDRIVETARALVDABGLAAVSTRRLAELGVSGP
SLYNHFKTEILEAVDSVSAQVDSMEEDGEMWETALRDWASVYTRALRDPNIVP
VLAHGPPRRPALHLADAVYGMVRAGWPAQAQISICALMRVFMVMSALGSPAGGFVD
DAGAYDPADYVPHLQAHLLAEQKEIKDERAFETGLTLLDGLAQQAQVAQVQV"
2889..3029
/misc_feature
/genes="SC01702"

```

```

/notes="Pfam match to entry PF00440 tetr, Bacterial
regulatory proteins, tetr family, score 45.00, E-value
9.2e-11."
2925..3017
/genes="SC01702"
/notes="PS01081 Bacterial regulatory proteins, tetr family
signature."
3555..4268
/genes="SC01703"
/notes="synonym: SCI30A.24"
3555..4268
/genes="SC01703"
/notes="SCI30A.24, possible transcriptional regulator, len:
237aa; similar to many eg. SW:MERR STRLI probable mercury
resistance operon repressor from Streptomyces lividans
(125 aa) fasta scores; opt: 144, z-score: 183.3, E():
0.007, (38.2% identity in 89 aa overlap). Contains Pfam
match to entry PF01022 HTH 5, Bacterial regulatory
protein, arSR family. Contains possible helix-turn-helix
motif (+3.92 SD) 37-58aa."
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB46801.1"
/db_xref="GI:5441777"
/db_xref="GOA:Q9S249"
/db_xref="SPTREMBL:Q9S249"

Alignment Scores:
Pred. No.: 2.02e+04 Length: 283100
Score: 97.50 Matches: 28
Percent Similarity: 33.01% Conservative: 6
Best Local Similarity: 27.18% Mismatches: 20
Query Match: 21.38% Indels: 49
DB: 1 Gaps: 6

US-09-972-032-2 (1-79) x SCO939110 (1-283100)
QY 3 GlyArgProArgAgg-----ValSerAlaGlyCysGlyPheAla 15
Db 169946 GGGCGGTCGCGCGAGCTGGGGCCCGCTGCGGAGCAGCGCGGTTGTCGTCG 170005
QY 16 AspaLahistp-----TprThrglyLeuGlyGluGly 29
Db 170006 CCTCGCACTGGCTCGCGAGCGGGTGGAGCGGTGACCCCGCGAGCGCGCG 170065
QY 20 -----ThrGlyLeu-----TprThrglyLeuGlyGluGly 29
Db 170086 CCCGGGATGGGGTACCGCGGTGAGCGGTGCTGCTGCGCGCGG 170119
QY 30 GlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArg 49
Db 170120 GCGCGGGCCGTCGGGGCCCTCGCGGGGGTGGCGCCCGCGCGCGCG 170176
QY 50 TrpProArgSerAlaSer-----Arg 56
Db 170177 TGGCGGAGAGCGCTTACGACCGCGGACTCGATGATCACCAGTTCGAGTTCGCGGCGC 170236
QY 57 TrpProTrp 59
Db 170237 TGGCCCTGG 170245

RESULT 15
AC113007/c
LOCUS
DEFINITION Mus musculus clone RP23-349H1.3, WORKING DRAFT SEQUENCE, 11
unordered pieces.
ACCESSION AC113007
VERSION HTG: HTGS_PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```

```

Consensus quality: 191584 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 192012; sum-of-contrigs
Quality coverage: 6.4 in Q20 bases; agarose-fp
Quality coverage: 7.0 in Q20 bases; sum-of-contrigs
-----
** NOTE: This is a 'working draft' sequence. It currently
** consists of 11 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contrigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.
1
56113: contig of 56113 bp in length
56114: gap of 100 bp
56115: gap of 100 bp
56116: gap of 100 bp
56117: gap of 100 bp
56118: gap of 100 bp
56119: gap of 100 bp
56120: gap of 100 bp
56121: gap of 100 bp
56122: gap of 100 bp
56123: gap of 100 bp
56124: gap of 100 bp
56125: gap of 100 bp
56126: gap of 100 bp
56127: gap of 100 bp
56128: gap of 100 bp
56129: gap of 100 bp
56130: gap of 100 bp
56131: gap of 100 bp
56132: gap of 100 bp
56133: gap of 100 bp
56134: gap of 100 bp
56135: gap of 100 bp
56136: gap of 100 bp
56137: gap of 100 bp
56138: gap of 100 bp
56139: gap of 100 bp
56140: gap of 100 bp
56141: gap of 100 bp
56142: gap of 100 bp
56143: gap of 100 bp
56144: gap of 100 bp
56145: gap of 100 bp
56146: gap of 100 bp
56147: gap of 100 bp
56148: gap of 100 bp
56149: gap of 100 bp
56150: gap of 100 bp
56151: gap of 100 bp
56152: gap of 100 bp
56153: gap of 100 bp
56154: gap of 100 bp
56155: gap of 100 bp
56156: gap of 100 bp
56157: gap of 100 bp
56158: gap of 100 bp
56159: gap of 100 bp
56160: gap of 100 bp
56161: gap of 100 bp
56162: gap of 100 bp
56163: gap of 100 bp
56164: gap of 100 bp
56165: gap of 100 bp
56166: gap of 100 bp
56167: gap of 100 bp
56168: gap of 100 bp
56169: gap of 100 bp
56170: gap of 100 bp
56171: gap of 100 bp
56172: gap of 100 bp
56173: gap of 100 bp
56174: gap of 100 bp
56175: gap of 100 bp
56176: gap of 100 bp
56177: gap of 100 bp
56178: gap of 100 bp
56179: gap of 100 bp
56180: gap of 100 bp
56181: gap of 100 bp
56182: gap of 100 bp
56183: gap of 100 bp
56184: gap of 100 bp
56185: gap of 100 bp
56186: gap of 100 bp
56187: gap of 100 bp
56188: gap of 100 bp
56189: gap of 100 bp
56190: gap of 100 bp
56191: gap of 100 bp
56192: gap of 100 bp
56193: gap of 100 bp
56194: gap of 100 bp
56195: gap of 100 bp
56196: gap of 100 bp
56197: gap of 100 bp
56198: gap of 100 bp
56199: gap of 100 bp
56200: gap of 100 bp
56201: gap of 100 bp
56202: gap of 100 bp
56203: gap of 100 bp
56204: gap of 100 bp
56205: gap of 100 bp
56206: gap of 100 bp
56207: gap of 100 bp
56208: gap of 100 bp
56209: gap of 100 bp
56210: gap of 100 bp
56211: gap of 100 bp
56212: gap of 100 bp
56213: gap of 100 bp
56214: gap of 100 bp
56215: gap of 100 bp
56216: gap of 100 bp
56217: gap of 100 bp
56218: gap of 100 bp
56219: gap of 100 bp
56220: gap of 100 bp
56221: gap of 100 bp
56222: gap of 100 bp
56223: gap of 100 bp
56224: gap of 100 bp
56225: gap of 100 bp
56226: gap of 100 bp
56227: gap of 100 bp
56228: gap of 100 bp
56229: gap of 100 bp
56230: gap of 100 bp
56231: gap of 100 bp
56232: gap of 100 bp
56233: gap of 100 bp
56234: gap of 100 bp
56235: gap of 100 bp
56236: gap of 100 bp
56237: gap of 100 bp
56238: gap of 100 bp
56239: gap of 100 bp
56240: gap of 100 bp
56241: gap of 100 bp
56242: gap of 100 bp
56243: gap of 100 bp
56244: gap of 100 bp
56245: gap of 100 bp
56246: gap of 100 bp
56247: gap of 100 bp
56248: gap of 100 bp
56249: gap of 100 bp
56250: gap of 100 bp
56251: gap of 100 bp
56252: gap of 100 bp
56253: gap of 100 bp
56254: gap of 100 bp
56255: gap of 100 bp
56256: gap of 100 bp
56257: gap of 100 bp
56258: gap of 100 bp
56259: gap of 100 bp
56260: gap of 100 bp
56261: gap of 100 bp
56262: gap of 100 bp
56263: gap of 100 bp
56264: gap of 100 bp
56265: gap of 100 bp
56266: gap of 100 bp
56267: gap of 100 bp
56268: gap of 100 bp
56269: gap of 100 bp
56270: gap of 100 bp
56271: gap of 100 bp
56272: gap of 100 bp
56273: gap of 100 bp
56274: gap of 100 bp
56275: gap of 100 bp
56276: gap of 100 bp
56277: gap of 100 bp
56278: gap of 100 bp
56279: gap of 100 bp
56280: gap of 100 bp
56281: gap of 100 bp
56282: gap of 100 bp
56283: gap of 100 bp
56284: gap of 100 bp
56285: gap of 100 bp
56286: gap of 100 bp
56287: gap of 100 bp
56288: gap of 100 bp
56289: gap of 100 bp
56290: gap of 100 bp
56291: gap of 100 bp
56292: gap of 100 bp
56293: gap of 100 bp
56294: gap of 100 bp
56295: gap of 100 bp
56296: gap of 100 bp
56297: gap of 100 bp
56298: gap of 100 bp
56299: gap of 100 bp
56300: gap of 100 bp
56301: gap of 100 bp
56302: gap of 100 bp
56303: gap of 100 bp
56304: gap of 100 bp
56305: gap of 100 bp
56306: gap of 100 bp
56307: gap of 100 bp
56308: gap of 100 bp
56309: gap of 100 bp
56310: gap of 100 bp
56311: gap of 100 bp
56312: gap of 100 bp
56313: gap of 100 bp
56314: gap of 100 bp
56315: gap of 100 bp
56316: gap of 100 bp
56317: gap of 100 bp
56318: gap of 100 bp
56319: gap of 100 bp
56320: gap of 100 bp
56321: gap of 100 bp
56322: gap of 100 bp
56323: gap of 100 bp
56324: gap of 100 bp
56325: gap of 100 bp
56326: gap of 100 bp
56327: gap of 100 bp
56328: gap of 100 bp
56329: gap of 100 bp
56330: gap of 100 bp
56331: gap of 100 bp
56332: gap of 100 bp
56333: gap of 100 bp
56334: gap of 100 bp
56335: gap of 100 bp
56336: gap of 100 bp
56337: gap of 100 bp
56338: gap of 100 bp
56339: gap of 100 bp
56340: gap of 100 bp
56341: gap of 100 bp
56342: gap of 100 bp
56343: gap of 100 bp
56344: gap of 100 bp
56345: gap of 100 bp
56346: gap of 100 bp
56347: gap of 100 bp
56348: gap of 100 bp
56349: gap of 100 bp
56350: gap of 100 bp
56351: gap of 100 bp
56352: gap of 100 bp
56353: gap of 100 bp
56354: gap of 100 bp
56355: gap of 100 bp
56356: gap of 100 bp
56357: gap of 100 bp
56358: gap of 100 bp
56359: gap of 100 bp
56360: gap of 100 bp
56361: gap of 100 bp
56362: gap of 100 bp
56363: gap of 100 bp
56364: gap of 100 bp
56365: gap of 100 bp
56366: gap of 100 bp
56367: gap of 100 bp
56368: gap of 100 bp
56369: gap of 100 bp
56370: gap of 100 bp
56371: gap of 100 bp
56372: gap of 100 bp
56373: gap of 100 bp
56374: gap of 100 bp
56375: gap of 100 bp
56376: gap of 100 bp
56377: gap of 100 bp
56378: gap of 100 bp
56379: gap of 100 bp
56380: gap of 100 bp
56381: gap of 100 bp
56382: gap of 100 bp
56383: gap of 100 bp
56384: gap of 100 bp
56385: gap of 100 bp
56386: gap of 100 bp
56387: gap of 100 bp
56388: gap of 100 bp
56389: gap of 100 bp
56390: gap of 100 bp
56
```

Alignment Scores:  
Pred. No.:  
Score:

193012  
27

Percent Similarity: 35.11% Conservative: 6  
Best Local Similarity: 28.72% Mismatches: 41  
Query Match: 21.27% Indels: 20  
DB: 2 Gaps: 3

US-09-972-032-2 (1-79) x AC113007 (1-193012)

Qy 5 ProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThr 24  
Db 53982 CCGCGGCGAGTCGGGGCCCCCGGGGCGAGCTGCACGGCCACGTGAGAGGTGAGTGCGG 53923  
Qy 25 -----GlyLeuGlyGluGlyGlyIle 34  
Db 53922 GCGGGCGCGCATGTTCTCGGCTGCGGGCGGCTGCGGGCGCTCGGGTGGCGCTCGG 53863  
Qy 35 GlyProGluGlyGlnAlaSer-----ProThrProAsp 45  
Db 53862 GGGTCCCGGGCGCGCACCGGATCGGGCGGTGCGGCTACCGGCCCTGTCTCCGCGCTCCC 53803  
Qy 46 CysAlaSerArgTrpProArgSerAlaSerArg--TipProTrpSerAlaGlyLeuThr 64  
Db 53802 TCGGGCCCCCGTGGGGCGCGGACAGGACTAGTGGCCGGTGAGCCGGCGGCGAGGTGC 53743  
Qy 65 ValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArg 78  
Db 53742 GTGGCCGAGTGCTGTGAACCTCGGTGACCTCTGCGAGGGTCGA 53701

Search completed: July 10, 2004, 17:50:05  
Job time : 3147 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 10, 2004, 14:37:10 ; Search time 538 Seconds  
(without alignments)  
8968.670 Million cell updates/sec

Title: US-09-972-032-1  
Perfect score: 990  
Sequence: 1 ggaattgtcttcgagccaa.....aaaatgagcgccgaagt 990

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 3183909 seqs, 2436941669 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published Applications NA:\*
- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
  - 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
  - 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
  - 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
  - 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
  - 17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
  - 18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
  - 19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	13	US-09-972-032-1
2	80	8.1	404	10	Sequence 1, Appli
3	55	5.6	114615	13	Sequence 1098, Ap
4	54	5.5	174448	13	Sequence 676, App
5	51	5.2	149480	10	Sequence 148, App
6	51	5.2	149480	10	Sequence 284, App
7	51	5.2	149480	12	Sequence 285, App
8	50	5.1	215221	13	Sequence 232, App
9	49	4.9	738	13	Sequence 1360, Ap
10	49	4.9	827	13	Sequence 23812, A
11	49	4.9	827	13	Sequence 23812, A
12	49	4.9	827	13	Sequence 167845,
13	49	4.9	827	13	Sequence 167846,
14	49	4.9	827	16	Sequence 167845,

Sequence 167846,  
Sequence 167847,  
Sequence 32, Appl  
Sequence 1463, Ap  
Sequence 264705,  
Sequence 264706,  
Sequence 264705,  
Sequence 264706,  
Sequence 1762, Ap  
Sequence 1795, Ap  
Sequence 1451, Ap  
Sequence 4, Appl  
Sequence 46, Appl  
Sequence 46, Appl  
Sequence 43293, A  
Sequence 43293, A  
Sequence 17785, A  
Sequence 17785, A  
Sequence 46783, A  
Sequence 46783, A  
Sequence 73773, A  
Sequence 73773, A  
Sequence 313211,  
Sequence 313211,  
Sequence 73773, A  
Sequence 313211,  
Sequence 254288,  
Sequence 254289,  
Sequence 254288,  
Sequence 254289,  
Sequence 330, App  
Sequence 127, App  
Sequence 1356, App

US-10-027-632-167846  
US-10-027-632-167847  
US-10-081-327-32  
US-10-292-798-1463  
US-10-027-632-264705  
US-10-027-632-264706  
US-10-027-632-264705  
US-10-027-632-264706  
US-10-087-192-1762  
US-10-087-192-1762  
US-10-292-798-1451  
US-10-105-637-4  
US-10-034-650-46  
US-10-450-826-46  
US-10-085-783A-43293  
US-10-242-535A-43293  
US-10-027-632-17785  
US-10-027-632-17785  
US-10-027-632-46783  
US-10-027-632-46783  
US-10-027-632-73773  
US-10-027-632-73773  
US-10-027-632-313211  
US-10-027-632-73773  
US-10-027-632-313211  
US-10-027-632-254288  
US-10-027-632-254289  
US-10-027-632-254288  
US-10-027-632-254289  
US-10-296-115-330  
US-10-108-260A-127  
US-10-104-047-1356

ALIGNMENTS

RESULT 1  
US-09-972-032-1  
Sequence 1, Application US/09972032  
Publication No. US20020086361A1  
GENERAL INFORMATION:  
APPLICANT: Case Western Reserve University  
APPLICANT: Montano, Monica  
APPLICANT: Sutton, Amelia  
TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology  
FILE REFERENCE: 27708/04003  
CURRENT APPLICATION NUMBER: US/09/972,032  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/238,190  
PRIOR FILING DATE: 2000-10-05  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 1  
LENGTH: 990  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-972-032-1

Query Match	100.0%	Score	990	DB	13	Length	990
Best Local Similarity	100.0%	Pred. No.	0				
Matches	990	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	1	GGAATGTTCTCGAGGCCAAGAAATCGGAGGATCTCTGACCTCGTATCGCCCG	60				
DB	1	GGAATGTTCTCGAGGCCAAGAAATCGGAGGATCTCTGACCTCGTATCGCCCG	60				
QY	61	CCTCAGCTTCCCAAGTCTGGGATTACAGGCATGAGCAGCGCCGCCCGCCGAC	120				
DB	61	CCTCAGCTTCCCAAGTCTGGGATTACAGGCATGAGCAGCGCCGCCCGCCGAC	120				
QY	121	CTAGTTTAAAGGCCCTGCTGTGTGCTGCGCGCTCCACGCTCCAGTCTG	180				
DB	121	CTAGTTTAAAGGCCCTGCTGTGTGCTGCGCGCTCCACGCTCCAGTCTG	180				

```
QY 181 GCGGCTCAGTCCCGCGTTCCTCCATGTTGTGGAGACCGCGTTCGCTAAGCGCTGGATGTGG 240
Db 181 GCGGCTCAGTCCCGCGTTCCTCCATGTTGTGGAGACCGCGTTCGCTAAGCGCTGGATGTGG 240
QY 241 CTTGCGTGTGATGCAATTTGACCGGCTCTGAGCTAGGCTAGGGAAGGCGAGGCGCGG 300
Db 241 CTTGCGTGTGATGCAATTTGACCGGCTCTGAGCTAGGGAAGGCGAGGCGCGG 300
QY 301 AATTGGCCCCAGGGCCAGAGCCCTCGCGACCCCGACTGCGGCTCCCGTGGCCCGCGAG 360
Db 301 AATTGGCCCCAGGGCCAGAGCCCTCGCGACCCCGACTGCGGCTCCCGTGGCCCGCGAG 360
QY 361 CGCTCCCGGTGGCCCTGGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGA 420
Db 361 CGCTCCCGGTGGCCCTGGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGA 420
QY 421 GCTGTGATGGGCGTGGCTTAAGCGCGTGGTTTGGTTAGCATTTGGCCAGCGGACTTAAG 480
Db 421 GCTGTGATGGGCGTGGCTTAAGCGCGTGGTTTGGTTAGCATTTGGCCAGCGGACTTAAG 480
QY 481 TGTGTCTCTGAAGAGCATGGAATTTAGTCTGAGGGTCTGGAAGAGTATCCCGCGCC 540
Db 481 TGTGTCTCTGAAGAGCATGGAATTTAGTCTGAGGGTCTGGAAGAGTATCCCGCGCC 540
QY 541 CACCATCAATGGCGTTAGTCTAGGAAGCGGTGTGGGTGGGCGCTTAGGGCGAGGCG 600
Db 541 CACCATCAATGGCGTTAGTCTAGGAAGCGGTGTGGGTGGGCGCTTAGGGCGAGGCG 600
QY 601 CAGACATACCCCAAGTGGTTGATTTATPACCGCAAGGGCTTGGATCGAATCCCGCAAA 660
Db 601 CAGACATACCCCAAGTGGTTGATTTATPACCGCAAGGGCTTGGATCGAATCCCGCAAA 660
QY 661 GACACTGGAGGCTGTGTGCTCAGAGGCGCGGCAATCCAGTGTGTGGGCTTTAC 720
Db 661 GACACTGGAGGCTGTGTGCTCAGAGGCGCGGCAATCCAGTGTGTGGGCTTTAC 720
QY 721 AGGAAAGAGCTCCACCTTCTGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780
Db 721 AGGAAAGAGCTCCACCTTCTGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780
QY 781 GAGTGGGCGCGGCGAGATGTCGCCAGTACAAAGCTGATTTGACCTGGGCGCTCTG 840
Db 781 GAGTGGGCGCGGCGAGATGTCGCCAGTACAAAGCTGATTTGACCTGGGCGCTCTG 840
QY 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTGCTCCCGTTGGCTGCCTTCAT 900
Db 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTGCTCCCGTTGGCTGCCTTCAT 900
QY 901 CCACCTCTCAGTCTCTGCTTTCAGATTAATTCAGATCTGTGGTGCAGGCAAAAAA 960
Db 901 CCACCTCTCAGTCTCTGCTTTCAGATTAATTCAGATCTGTGGTGCAGGCAAAAAA 960
QY 961 AAAAAAATAAATAAGCGGCGCAAGTT 990
Db 961 AAAAAAATAAATAAGCGGCGCAAGTT 990
```

## RESULT 2

```
US-09-764-891-1098
; Sequence 1098, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1098
; LENGTH: 404
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (378)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (380)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1098
```

```
Query Match 8.1%; Score 80; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 872 CAAAGTCTCTGCCGTTGGCTTTCATCCACTCTCTCACTTCTCTGCTTTCAGAGTAA 931
Db 8 CAAAGTCTCTGCCGTTGGCTTTCATCCACTCTCTCACTTCTCTGCTTTCAGAGTAA 67
QY 932 AATTGCAAGATCTCTGTGTC 951
Db 68 AATTGCAAGATCTCTGTGTC 87
```

## RESULT 3

```
US-10-087-192-676
; Sequence 676, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 676
; LENGTH: 114615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(114615)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-676
```

```
Query Match 5.6%; Score 55; DB 13; Length 114615;
Best Local Similarity 100.0%; Pred. No. 8.4e-18;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 38 CTCCTGACCTCGTGTATCCGCGGCTCAGCTTCCCAAAGTGTGGGATTACAGGC 92
Db 57568 CTCCTGACCTCGTGTATCCGCGGCTCAGCTTCCCAAAGTGTGGGATTACAGGC 57622
```

## RESULT 4

```
US-10-087-192-148/c
; Sequence 148, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
```



```
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 174448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(174448)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-148

Query Match          5.2%; Score 54; DB 13; Length 174448;
Best Local Similarity 100.0%; Pred. No. 2.6e-17; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 0;

QY 49 GTGATCGCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACC 102
Db 80771 GTGATCGCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACC 80718

RESULT 5
US-09-873-367C-284
; Sequence 284, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meana
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 284
; LENGTH: 149480
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-284

Query Match          5.2%; Score 51; DB 10; Length 149480;
Best Local Similarity 100.0%; Pred. No. 8.8e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 57 CCGCCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACCGGCC 107
Db 76386 CCGCCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACCGGCC 76436

RESULT 7
US-09-968-007A-232
; Sequence 232, Application US/09968007A
; Publication No. US20040115825A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-71
; CURRENT APPLICATION NUMBER: US/09/968,007A
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,172
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,173
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,278
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,294
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,295
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,316
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 1001
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 232
; LENGTH: 149480
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-968-007A-232

Query Match          5.2%; Score 51; DB 12; Length 149480;
Best Local Similarity 100.0%; Pred. No. 8.8e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 57 CCGCCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACCGGCC 107
Db 76386 CCGCCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGCCACCGGCC 107
```

```
Db 76386 CCGCCTCAGCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCC 76436
|||||
RESULT 8
US-10-087-192-1360
; Sequence 1360, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1360
; LENGTH: 215221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1360

Query Match 5.1%; Score 50; DB 13; Length 215221;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 83
|||||
Db 49208 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 49257
|||||

RESULT 9
US-10-027-632-23812
; Sequence 23812, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 13; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 82
|||||
Db 368 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 416
|||||

RESULT 11
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 13; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 34 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 82
|||||
Db 368 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 416
|||||

RESULT 10
US-10-027-632-23812
; Sequence 23812, Application US/10027632
; Publication No. US20020204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 16; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 82
|||||
Db 368 CGATCTCTGACCTCGATCCGCGCCGCTCAGCTTCCCAAAGTGTGG 416
|||||

RESULT 11
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167845
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167845

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 12
US-10-027-632-167846/c
; Sequence 167846, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167846
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167846

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 13
US-10-027-632-167847/c
; Sequence 167847, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167847
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167847

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 14
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167845
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167845

Query Match
Best Local Similarity 4.9%; Score 49; DB 16; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 15
US-10-027-632-167846/c
; Sequence 167846, Application US/10027632
```

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167846
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167846

Query Match          4.9%; Score 49; DB 16; Length 827;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      56 GCCGCTCAGCTCCCAAAGTGTGGGATTACAGGCATGAGCCACGC 104
        |||||||
Db      704 GCCGCTCAGCTCCCAAAGTGTGGGATTACAGGCATGAGCCACGC 656
```

Search completed: July 10, 2004, 16:51:03  
Job time : 544 secs

29: em\_vt:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	79	100.0	0	9	AC135044	
2	79	100.0	195476	2	AC135044	
3	11	13.9	195476	1	AC135044	
C 4	9	11.4	195476	1	AC135044	
C 5	9	11.4	195476	6	AC135044	
6	9	11.4	195476	3	AC135044	
7	9	11.4	195476	3	AC135044	
8	9	11.4	195476	2	AC135044	
9	9	11.4	195476	2	AC135044	
10	9	11.4	195476	2	AC135044	
11	9	11.4	195476	9	AC135044	
C 12	9	11.4	195476	2	AC135044	
C 13	9	11.4	195476	2	AC135044	
C 14	9	11.4	195476	9	AC135044	
15	9	11.4	195476	2	AC135044	
16	9	11.4	195476	2	AC135044	
17	9	11.4	195476	2	AC135044	
18	9	11.4	195476	2	AC135044	
19	9	11.4	195476	9	AC135044	
C 20	9	11.4	195476	10	AC135044	
C 21	9	11.4	195280	9	AC073869	
C 22	9	11.4	195280	2	AC073869	
C 23	9	11.4	195280	2	AC073869	
C 24	9	11.4	195280	1	AC073869	
25	8	10.1	195280	9	AC073869	
26	8	10.1	195280	9	AC073869	
C 27	8	10.1	195280	6	AC073869	
28	8	10.1	195280	1	AC073869	
C 29	8	10.1	195280	6	AC073869	
C 30	8	10.1	195280	6	AC073869	
31	8	10.1	195280	10	AC073869	
C 32	8	10.1	195280	9	AC073869	
33	8	10.1	195280	4	AC073869	
C 34	8	10.1	195280	9	AC073869	
35	8	10.1	195280	9	AC073869	
C 36	8	10.1	195280	6	AC073869	
37	8	10.1	195280	9	AC073869	
38	8	10.1	195280	9	AC073869	
39	8	10.1	195280	9	AC073869	
40	8	10.1	195280	9	AC073869	
41	8	10.1	195280	9	AC073869	
42	8	10.1	195280	9	AC073869	
43	8	10.1	195280	9	AC073869	
44	8	10.1	195280	9	AC073869	
45	8	10.1	195280	9	AC073869	

ALIGNMENTS

RESULT 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: July 10, 2004, 18:59:03 ; Search time 2822 Seconds  
(without alignments)  
1213.359 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 79  
Sequence: 1 MCGRRPRVSAGCGFADAHWT.....SAGLTVRDRPOLGELCMGRG 79

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 6933665

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09972032/runat\_06072004\_121534\_8872/app\_query.fasta\_1.263  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972032@cgn.1.1.3731 @runat\_06072004\_121534\_8872 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_m.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*



AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>

Project Information  
 Center Project Name: 809609  
 Center clone name: CITB-El\_2551B20

-----Summary Statistics  
 Consensus quality: 171229 bases at least Q40  
 Consensus quality: 182638 bases at least Q30  
 Consensus quality: 18095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-contigs estimation  
 Quality coverage: 6.7 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

Alignment Scores:  
 Pred. No.: 46.1 Length: 195476  
 Score: 11.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 13.92% Indels: 0  
 DB: 1 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
 QY 32 GlyGlyIleGlyPro 36  
 Db 98566 GGTGGATAGCCCC 98580

RESULT 4  
 AC135044/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
 SEQUENCE, 22 unordered pieces.  
 ACSSSION  
 AC135044  
 VERSION AC135044.1 GI:23505535  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>

-----Project Information  
 Center Project Name: 809609  
 Center clone name: CITB-El\_2551B20

-----Summary Statistics  
 Consensus quality: 171229 bases at least Q40  
 Consensus quality: 182638 bases at least Q30  
 Consensus quality: 18095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-contigs estimation  
 Quality coverage: 6.7 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

Alignment Scores:  
 Pred. No.: 45.3 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 1 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
 QY 32 GlyGlyIleGlyPro 36  
 Db 162005 GGGGTATTGGCCG 161991

RESULT 5  
 AC135044/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
 SEQUENCE, 22 unordered pieces.  
 ACSSSION  
 AC135044  
 VERSION AC135044.1 GI:23505535  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>

-----Project Information  
 Center Project Name: 809609  
 Center clone name: CITB-El\_2551B20

-----Summary Statistics  
 Consensus quality: 171229 bases at least Q40  
 Consensus quality: 182638 bases at least Q30

Consensus quality: 188095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-contigs estimation  
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 22 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 be preserved.

Alignment Scores:  
 Pred. No.: 45.3 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
 QY 32 GlyGlyIleGlyPro 36  
 Db 162005 GGGGGTATTGGCGG 161991

RESULT 6  
 AC135044  
 LOCUS  
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT

ACCESSION AC135044  
 VERSION AC135044.1 GI:23505535

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 JOURNAL

REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 16  
 JOURNAL

REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission

COMMENT  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov

Project Information  
 Center Project Name: 809609  
 Center clone name: CITB-EL\_2551B20

Summary Statistics  
 Consensus quality: 171229 bases at least Q40  
 Consensus quality: 182638 bases at least Q30  
 Consensus quality: 188095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-contigs estimation  
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 22 contigs. The true order of the pieces  
 is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 be preserved.

Alignment Scores:  
 Pred. No.: 198 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
 QY 32 GlyGlyIleGlyPro 36  
 Db 98566 GGTGGATAGGCCCC 98580

RESULT 7  
 AC135044  
 LOCUS  
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT

ACCESSION AC135044  
 VERSION AC135044.1 GI:23505535

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 JOURNAL

REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 16  
 JOURNAL

REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission

COMMENT  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov

Project Information  
 Center Project Name: 809609  
 Center clone name: CITB-EL\_2551B20

Summary Statistics  
 Consensus quality: 171229 bases at least Q40  
 Consensus quality: 182638 bases at least Q30  
 Consensus quality: 188095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-contigs estimation  
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 22 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 be preserved.

Alignment Scores:  
 Pred. No.: 198 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 3 Gaps: 0



\* 2469 2568: gap of unknown

Alignment Scores:  
Pred. No.: 206 Length: 195476  
Score: 9.00 Matches: 5  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
QY 32 GlyGlyIleGlyPro 36  
Db 98566 GGTGGATAGGCCCC 98580

RESULT 8  
AC135044  
LOCUS  
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 809609  
Center clone name: CITB-El\_2551B20

-----  
Summary Statistics  
Consensus quality: 171229 bases at least Q40  
Consensus quality: 182638 bases at least Q30  
Consensus quality: 188095 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Alignment Scores:  
Pred. No.: 1  
Score: 1203 Length: 1202 bp in length  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
QY 32 GlyGlyIleGlyPro 36  
Db 98566 GGTGGATAGGCCCC 98580

RESULT 9  
AC135044  
LOCUS  
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 809609  
Center clone name: CITB-El\_2551B20

-----  
Summary Statistics  
Consensus quality: 171229 bases at least Q40  
Consensus quality: 182638 bases at least Q30  
Consensus quality: 188095 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Alignment Scores:  
Pred. No.: 1  
Score: 1203 Length: 1202 bp in length  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
QY 32 GlyGlyIleGlyPro 36  
Db 98566 GGTGGATAGGCCCC 98580

RESULT 9  
AC135044  
LOCUS  
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 809609  
Center clone name: CITB-El\_2551B20

-----  
Summary Statistics  
Consensus quality: 171229 bases at least Q40  
Consensus quality: 182638 bases at least Q30  
Consensus quality: 188095 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Alignment Scores:  
Pred. No.: 1  
Score: 1203 Length: 1202 bp in length  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
QY 32 GlyGlyIleGlyPro 36  
Db 98566 GGTGGATAGGCCCC 98580

RESULT 9  
AC135044  
LOCUS  
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 809609  
Center clone name: CITB-El\_2551B20

-----  
Summary Statistics  
Consensus quality: 171229 bases at least Q40  
Consensus quality: 182638 bases at least Q30  
Consensus quality: 188095 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Alignment Scores:  
Pred. No.: 1  
Score: 1203 Length: 1202 bp in length  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)  
QY 32 GlyGlyIleGlyPro 36  
Db 98566 GGTGGATAGGCCCC 98580

RESULT 9  
AC135044  
LOCUS  
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 809609  
Center clone name: CITB-El\_2551B20

-----  
Summary Statistics  
Consensus quality: 171229 bases at least Q40  
Consensus quality: 182638 bases at least Q30  
Consensus quality: 188095 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Alignment Scores:  
Pred. No.: 1  
Score: 1203 Length: 1202 bp in length  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.39% Indels: 0  
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044



REFERENCE	2 (bases 1 to 195476)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	-----Genome Center Center: Joint Genome Institute Center Code: JGI web site: <a href="http://www.jgi.doe.gov">http://www.jgi.doe.gov</a>

```
-----
Project Information
Center Project Name: 809609
Center clone name: CITB-EI 2551B20
```

Summary Statistics

Consensus quality:	171229 bases at least Q40
Consensus quality:	182638 bases at least Q30
Consensus quality:	188095 bases at least Q20
Estimated insert size:	170000; agarose-fp estimation
Estimated insert size:	193376; sum-of-contigs estimation
Quality coverage:	7.63 in Q20 bases; agarose-fp estimation
Quality coverage:	5.7 in Q20 bases; sum-of-contigs estimation.

\* NOTE: this is a 'working draft' sequence. it currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

\* 1 1202: contig of 1202 bp in length  
 \* 1203 1302: gap of unknown length  
 \* 1303 2468: contig of 1166 bp in length  
 \* 2469 2568: gap of unknown

Alignment Scores:	1.97e+03	Length:	195476
Pred. No.:	9.00	Matches:	5
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	11.39%	Gaps:	0
DB:	2		

US-09-972-032-2 (1-79) x AC135044 (1-195476)

Qy 32 GlyGlyIleGlyPro 36  
Db 162005 GGGGGTATTGGGCCG 161991

RESULT 13	
AC135044/C	
LOCUS	
AC135044	195476 bp DNA linear HTG 05-OCT-2002
DEFINITION	Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT SEQUENCE. 22 unordered pieces.

SEQUENCE, 22 unaligned pieces.  
AC135044  
ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens (human);  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 195476)

REFERENCE 1 (bases 1 to 195476)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 16  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 195476)  
 AUTHORS DOE Joint Genome Institute.

ARRNO	
TITLE	DOE JOINT GENOME INSTITUTE.
JOURNAL	Direct Submission
COMMENT	Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA -----Genome Center Center: Joint Genome Institute Center Code: JGI

```

web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: CITB-El 2551B20

```

Summary Statistics	171229	bases at
Consensus quality:	182638	bases at
Consensus quality:	188095	bases at

```

Consensus quality: 10000 bases at least x20
Estimated insert size: 170000; agarose-1p estimation
Estimated insert size: 193376; sum-of-contigs estimation
Quality coverage: 7.63 in Q20 bases; agarose-1p estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 22 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence.
  * as soon as it is available and the accession number will
  * be preserved.

```

\* 1 1202: contig of 1202 bp in length  
 \* 1203 1302: gap of unknown length  
 \* 1303 2468: contig of 1166 bp in length  
 \* 2469 2568: gap of unknown

Alignment Scores:		
Pred. No.:	2.05e+03	Length:
Score:	9.00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	100.00%	Mismatches:
Query Match:	11.33%	Indels:
DB:	2	Gaps:
		195476

US-09-972-032-2 (1-79) x AC135044 (1-195476)

Ov 32 GlyGlyIleGlyPro 36

Db 162005 GGGGGTATTGGCCG 161991

RESULT 14  
AC135044/C

AC135044	195476 bp	DNA	linear	HTG 05-OCT-2002
LOCUS				
DEFINITION	Homo sapiens	chromosome 16	clone CTD-2551B20,	WORKING DRAFT
SEQUENCE	22	unordered	pieces.	

ACCESSION AC135044  
VERSION AC135044.1 GI:23505535  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)

SOURCE	ORGANISM
Homo sapiens (human)	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE  
1 (bases 1 to 195476)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished

**JOURNAL**      **published**

**REFERENCE**    **2 (bases 1 to 195476)**

**AUTHORS**       **DOE Joint Genome Institute.**

DOE JOINT GENOME INSTITUTE.  
 TITLE Direct Submission  
 JOURNAL  
 COMMENT  
 Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI

Project Information  
Center Project Name: 809609  
Center clone name: CITB-E1 2551B20

Summary Statistics  
Consensus quality: 171229 bases at least Q40

Consensus quality: 182638 bases at least Q30  
 Consensus quality: 188095 bases at least Q20  
 Estimated insert size: 170000; agarose-fp estimation  
 Estimated insert size: 193376; sum-of-ctnigs estimation  
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.7 in Q20 bases; sum-of-ctnigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently

\* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1202: contig of 1202 bp in length  
 \* 1203 1302: gap of unknown length  
 \* 1303 2468: contig of 1166 bp in length  
 \* 2469 2588: gap of unknown

## Alignment Scores:

Pred. No.: 2.18e+03 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 9 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

QY 32 GlyGlyIleGlyPro 36

Db 162005 GGGGATATGGGCGG 161991

## RESULT 15

AC135044

LOCUS

DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT

SEQUENCE, 22 unordered pieces.

AC135044

AC135044.1 GI:23505535

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 195476)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

REFERENCE 2 (bases 1 to 195476)

DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 809609

Center clone name: CITB-E1.2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40

Consensus quality: 182638 bases at least Q30

Consensus quality: 188095 bases at least Q20

Estimated insert size: 170000; agarose-fp estimation

Estimated insert size: 193376; sum-of-ctnigs estimation

Quality coverage: 7.63 in Q20 bases; agarose-fp estimation

Quality coverage: 6.7 in Q20 bases; sum-of-ctnigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 22 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1202: contig of 1202 bp in length  
 \* 1203 1302: gap of unknown length  
 \* 1303 2468: contig of 1166 bp in length  
 \* 2469 2588: gap of unknown

## Alignment Scores:

Pred. No.: 2.27e+03 Length: 195476  
 Score: 9.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

QY 32 GlyGlyIleGlyPro 36

Db 98566 GGTGGATAGGCCCC 98580

Search completed: July 10, 2004, 20:57:09

Job time : 3260 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 12, 2004, 19:22:20 / Search time 399 Seconds  
(without alignments)  
965.004 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 79

Sequence: 1 MGRPRVRSAGCGFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table:

OLIGO	
Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 3183909 seqs, 2436941669 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6353483

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121537_8984/app_query.fasta_1.263
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAFSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972032 @CGN_1_1_511 @runat_06072004_121537_8984
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

```

Database :

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
3:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
4:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
14:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
19:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	79	100.0	990	13	US-09-972-032-1	Sequence 1, Appli
2	11	13.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
3	8	10.1	250	17	US-10-437-963-31286	Sequence 31286, A
4	8	10.1	254	12	US-09-922-293-1599	Sequence 1599, Ap
5	8	10.1	507	15	US-10-156-761-6581	Sequence 6581, Ap
6	8	10.1	720	13	US-10-282-122A-11225	Sequence 11225, A
7	8	10.1	775	13	US-10-425-114-13766	Sequence 13766, A
8	8	10.1	777	16	US-10-369-493-35593	Sequence 35593, A
9	8	10.1	969	16	US-10-369-493-37767	Sequence 37767, A
10	8	10.1	975	17	US-10-437-963-38330	Sequence 38330, A
11	8	10.1	1007	13	US-10-425-114-21611	Sequence 21611, A
12	8	10.1	1035	13	US-10-282-122A-13444	Sequence 13444, A
13	8	10.1	1037	9	US-09-773-926-8	Sequence 7, Appli
14	8	10.1	1230	15	US-10-021-723A-7	Sequence 7, Appli
15	8	10.1	1239	13	US-10-282-122A-13867	Sequence 13867, A
16	8	10.1	1266	15	US-10-021-723A-5	Sequence 5, Appli
17	8	10.1	1275	13	US-10-425-114-35317	Sequence 35317, A
18	8	10.1	1396	17	US-10-468-456-1	Sequence 1, Appli
19	8	10.1	1437	9	US-09-815-242-7577	Sequence 7577, Ap
20	8	10.1	1536	13	US-10-282-122A-23591	Sequence 23591, A
21	8	10.1	1662	15	US-10-156-761-907	Sequence 907, App
22	8	10.1	1783	13	US-10-425-114-20795	Sequence 20795, A
23	8	10.1	1807	13	US-10-425-114-28103	Sequence 28103, A
24	8	10.1	1845	16	US-10-214-529-25	Sequence 25, Appl
25	8	10.1	1914	14	US-10-051-902-19	Sequence 19, Appl
26	8	10.1	1914	14	US-10-051-908-19	Sequence 19, Appl
27	8	10.1	2018	16	US-10-369-493-32746	Sequence 32746, A
28	8	10.1	2046	13	US-10-282-122A-14411	Sequence 14411, A
29	8	10.1	2053	17	US-10-437-963-4970	Sequence 4970, Ap
30	8	10.1	2097	15	US-10-156-761-1015	Sequence 1015, Ap
31	8	10.1	2169	16	US-10-094-749-814	Sequence 814, App
32	8	10.1	2172	15	US-10-156-761-5455	Sequence 5455, Ap
33	8	10.1	2664	15	US-10-132-350-19	Sequence 19, Appl
34	8	10.1	2751	17	US-10-437-963-38332	Sequence 38332, A
35	8	10.1	2855	13	US-10-425-114-35061	Sequence 35061, A
36	8	10.1	2929	13	US-09-940-227-8	Sequence 8, Appli
37	8	10.1	3003	13	US-10-425-114-34535	Sequence 34535, A
38	8	10.1	3039	16	US-10-108-260A-885	Sequence 885, App
39	8	10.1	3080	15	US-10-132-350-17	Sequence 17, Appl
40	8	10.1	4371	17	US-10-437-963-92083	Sequence 92083, A
41	8	10.1	5877	13	US-10-152-886-54	Sequence 54, Appl
42	8	10.1	9369	15	US-10-200-562-190	Sequence 190, App
43	8	10.1	9369	15	US-10-237-551-190	Sequence 190, App
44	8	10.1	9369	15	US-10-237-551-247	Sequence 247, App
45	8	10.1	12437	15	US-10-314-321A-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1

US-09-972-032-1

Sequence 1, Application US/09972032

Publication No. US20020086361A1

GENERAL INFORMATION:

APPLICANT: Case Western Reserve University

APPLICANT: Montano, Monica

APPLICANT: Sutton, Amelia

TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology

FILE REFERENCE: 27708/04003

CURRENT APPLICATION NUMBER: US/09/972,032

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: US 60/238,190

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 990

TYPE: DNA

ORGANISM: Homo sapiens

US-09-972-032-1

Alignment Scores:

```
Pred. No.: 1.68e-69 Length: 990
Score: 79.00 Matches: 79
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-09-972-032-1 (1-990)

Qy 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
Db 203 ATGTGTCGGAGACCGCTCGCTAAGCGCTGGATGTGGCTTCGCTGATGCACATTGGACC 262
Qy 21 GlyLeuTrpThrGlyLeuGlyGlnGlnGlyGlyGlnGlyGlyGlnGlnAla 40
Db 263 GGGCTCTGGACTGGCTAGGGAGGCGAGGAGGCGGGAATTGGGCCCCGAGGCGCCAGGCC 322
Qy 41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTpsr 60
Db 323 TCGCCGACCCCGACTGGCTCCCGTGGCGCCCGAGCGCCTCCCGGTGGCCCTGGACT 382
Qy 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyGlnLeuGlyArgGly 79
Db 383 GCAGGTCTTACCGTCCGAGATCTCCGCAACTGGCGAGCTGTGCATGGGGGTGGC 439

RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown.
US-10-156-761-1

Alignment Scores:
Pred. No.: 81.4 Length: 9025608
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.92% Indels: 0
DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-156-761-1 (1-9025608)

Qy 48 SerArgTrpProArgSerAlaSerArgTrpPro 58
Db 4452379 TCGCGTGGCCGAGAGTGCATCGGTGGCCA 4452411

RESULT 3
US-10-437-963-31286
; Sequence 31286, Application US/10437963
```

```
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 31286
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_35608C.1
US-10-437-963-31286

Alignment Scores:
Pred. No.: 25.2 Length: 250
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 17 Gaps: 0

US-09-972-032-2 (1-79) x US-10-437-963-31286 (1-250)

Qy 28 GluGlyGlnGlyGlyGlyGly 35
Db 130 GAAGGCAAGAGGAGGAGGATCGGT 153

RESULT 4
US-09-922-293-1599/c
; Sequence 1599, Application US/09922293
; Publication No. US20040123339A1
; GENERAL INFORMATION:
; APPLICANT: Comer, Timothy W.
; APPLICANT: Heck, Gregory R.
; APPLICANT: Liu, Jingsong
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 16517.254
; CURRENT APPLICATION NUMBER: US/09/922,293
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/067,000
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: US 60/069,472
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: US 60/071,479
; PRIOR FILING DATE: 1998-01-13
; PRIOR APPLICATION NUMBER: US 60/074,201
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,282
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,280
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,281
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,566
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/074,567
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/074,565
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/075,462
; PRIOR FILING DATE: 1998-02-19
```

;; PRIOR APPLICATION NUMBER: US 60/074,789  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/075,459  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/075,461  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/075,464  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/075,460  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/075,463  
;; PRIOR FILING DATE: 1998-02-19  
;; PRIOR APPLICATION NUMBER: US 60/077,231  
;; PRIOR FILING DATE: 1998-03-09  
;; PRIOR APPLICATION NUMBER: US 60/077,229  
;; PRIOR FILING DATE: 1998-03-09  
;; PRIOR APPLICATION NUMBER: US 60/077,230  
;; PRIOR FILING DATE: 1998-03-09  
;; PRIOR APPLICATION NUMBER: US 60/078,368  
;; PRIOR FILING DATE: 1998-03-18  
;; PRIOR APPLICATION NUMBER: US 60/080,844  
;; PRIOR FILING DATE: 1998-04-07  
;; PRIOR APPLICATION NUMBER: US 60/083,067  
;; PRIOR FILING DATE: 1998-04-27  
;; PRIOR APPLICATION NUMBER: US 60/083,386  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: US 60/083,387  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: US 60/083,388  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: US 60/083,389  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: US 60/085,224  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: US 60/085,223  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: US 60/085,222  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: US 60/085,533  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: US 60/086,186  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,187  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,185  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,184  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,183  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,188  
;; PRIOR FILING DATE: 1998-05-21  
;; PRIOR APPLICATION NUMBER: US 60/089,524  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: US 60/089,810  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,814  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,808  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,812  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,807  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,806  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,813  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,811  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/089,793  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: US 60/091,405

;; PRIOR FILING DATE: 1998-06-30  
;; PRIOR APPLICATION NUMBER: US 60/091,247  
;; PRIOR FILING DATE: 1998-06-30  
;; PRIOR APPLICATION NUMBER: US 60/099,667  
;; PRIOR FILING DATE: 1998-09-09  
;; PRIOR APPLICATION NUMBER: US 60/099,668  
;; PRIOR FILING DATE: 1998-09-09  
;; PRIOR APPLICATION NUMBER: US 60/099,670  
;; PRIOR FILING DATE: 1998-09-09  
;; PRIOR APPLICATION NUMBER: US 60/099,697  
;; PRIOR FILING DATE: 1998-09-09  
;; PRIOR APPLICATION NUMBER: US 60/100,674  
;; PRIOR FILING DATE: 1998-09-16  
;; PRIOR APPLICATION NUMBER: US 60/100,673  
;; PRIOR FILING DATE: 1998-09-16  
;; PRIOR APPLICATION NUMBER: US 60/100,672  
;; PRIOR FILING DATE: 1998-09-16  
;; PRIOR APPLICATION NUMBER: US 60/100,963  
;; PRIOR FILING DATE: 1998-09-17  
;; PRIOR APPLICATION NUMBER: US 60/101,131  
;; PRIOR FILING DATE: 1998-09-21  
;; PRIOR APPLICATION NUMBER: US 60/101,132  
;; PRIOR FILING DATE: 1998-09-21  
;; PRIOR APPLICATION NUMBER: US 60/101,130  
;; PRIOR FILING DATE: 1998-09-21  
;; PRIOR APPLICATION NUMBER: US 60/101,508  
;; PRIOR FILING DATE: 1998-09-22  
;; PRIOR APPLICATION NUMBER: US 60/101,344  
;; PRIOR FILING DATE: 1998-09-22  
;; PRIOR APPLICATION NUMBER: US 60/101,347  
;; PRIOR FILING DATE: 1998-09-22  
;; PRIOR APPLICATION NUMBER: US 60/104,128  
;; PRIOR FILING DATE: 1998-10-13  
;; PRIOR APPLICATION NUMBER: US 60/104,127  
;; PRIOR FILING DATE: 1998-10-13  
;; PRIOR APPLICATION NUMBER: US 60/104,124  
;; PRIOR FILING DATE: 1998-10-13  
;; PRIOR APPLICATION NUMBER: US 60/109,018  
;; PRIOR FILING DATE: 1998-11-18  
;; PRIOR APPLICATION NUMBER: US 60/108,996  
;; PRIOR FILING DATE: 1998-11-18  
;; PRIOR APPLICATION NUMBER: US 09/199,129  
;; PRIOR FILING DATE: 1998-11-24  
;; PRIOR APPLICATION NUMBER: US 09/210,297  
;; PRIOR FILING DATE: 1998-12-08  
;; PRIOR APPLICATION NUMBER: US 60/111,981  
;; PRIOR FILING DATE: 1998-12-11  
;; PRIOR APPLICATION NUMBER: US 60/113,224  
;; PRIOR FILING DATE: 1998-12-22  
;; PRIOR APPLICATION NUMBER: US 09/229,413  
;; PRIOR FILING DATE: 1999-01-12  
;; NUMBER OF SEQ ID NOS: 3853  
;; SEQ ID NO 1599  
;; LENGTH: 254  
;; TYPE: DNA  
;; ORGANISM: Zea mays  
US-09-922-293-1599  
  
Alignment Scores:  
Pred. No.: 25.5 Length: 254  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 12 Gaps: 0  
  
US-09-972-032-2 (1-79) x US-09-922-293-1599 (1-254)

Qy 52 ArgSerAlaSerArgTrpProTyr 59  
Db 31 CGATCAGGAGCCGTTGGCCATGG 8

## RESULT 5

US-10-156-761-6581  
; Sequence 6581, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6581  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(507)  
US-10-156-761-6581

Alignment Scores:  
Pred. No.: 43.3 Length: 507  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-156-761-6581 (1-507)

Qy 47 AlaSerArgTrpProArgSerAla 54  
Db 197 GCATCCCGTGGCCACGATCCGG 220

## RESULT 6

US-10-282-122A-11225/c  
; Sequence 11225, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 11225  
; LENGTH: 720  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-282-122A-11225

Alignment Scores:  
Pred. No.: 56.7 Length: 720  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-11225 (1-720)

Qy 47 AlaSerArgTrpProArgSerAla 54  
Db 107 GCGAGCCGCTGGCCGCTTCGGCC 84

## RESULT 7

US-10-425-114-13766/c  
; Sequence 13766, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 13766  
; LENGTH: 775  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB143-040-B9\_FLI  
US-10-425-114-13766

## Alignment Scores:

Pred. No.: 59.9 Length: 775  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 13 Gaps: 0



US-09-972-032-2 (1-79) x US-10-425-114-13766 (1-775)

QY 49 ArgTTPProArgSerAlaSerArg 56  
DB 312 CGCTGGCCGGGTCTGGCGTCGAGG 289

RESULT 8  
US-10-369-493-35593/C  
; Sequence 35593, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 35593  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Mesorhizobium loti  
US-10-369-493-35593

Alignment Scores: 777  
Pred. No.: 60.1 Length: 777  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 16 Gaps: 0

US-09-972-032-2 (1-79) x US-10-369-493-35593 (1-777)

QY 49 ArgTTPProArgSerAlaSerArg 56  
DB 649 CGATGGCTCGATCTGCGTCGCGC 626

RESULT 9  
US-10-369-493-37767  
; Sequence 37767, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 37767  
; LENGTH: 969  
; TYPE: DNA  
; ORGANISM: Pseudomonas fluorescens  
US-10-369-493-37767

Alignment Scores: 969  
Pred. No.: 71.1 Length: 969  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 10.13% Indels: 0  
DB: 16 Gaps: 0

US-09-972-032-2 (1-79) x US-10-369-493-37767 (1-969)

QY 47 AlaSerArgTTPProArgSerAla 54  
DB 617 GCAAGCCGCTGCCAAGGAGCGCT 640

RESULT 10  
US-10-437-963-38330  
; Sequence 38330, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 38330  
; LENGTH: 975  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_41977C.1  
US-10-437-963-38330

Alignment Scores: 975  
Pred. No.: 71.5 Length: 975  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 17 Gaps: 0

US-09-972-032-2 (1-79) x US-10-437-963-38330 (1-975)

QY 49 ArgTTPProArgSerAlaSerArg 56  
DB 329 AGGTGGCCTCGGTGCGCTAGCCGA 352

RESULT 11  
US-10-425-114-21611  
; Sequence 21611, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 21611  
; LENGTH: 1007  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3354-046-G9\_FLI

US-10-425-114-21611

Alignment Scores:  
Pred. No.: 73.3 Length: 1007  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-425-114-21611 (1-1007)

Qy 53 SerAlaSerArgTrpProTrpSer 60  
Db 309 TCAGCCTCGCGCTGGCCATGGAGC 332

RESULT 12

US-10-282-122A-13444/c  
; Sequence 13444, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13444

; LENGTH: 1035  
; TYPE: DNA

; ORGANISM: Burkholderia fungorum

US-10-282-122A-13444

Alignment Scores:  
Pred. No.: 74.8 Length: 1035  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-13444 (1-1035)

Qy 48 SerArgTrpProArgSerAlaSer 55  
Db 836 TCGAGATGCCGAGGTCGCCAGT 813

RESULT 13

US-09-773-926-8/c

; Sequence 8, Application US/09773926  
; Patent No. US20020042057A1

; GENERAL INFORMATION:  
; APPLICANT: KNOLL, RALPH

; TITLE OF INVENTION: MLP-GENE, NUCLEIC ACIDS, POLYPEPTIDES AND USE THEREOF

; FILE REFERENCE: SCH-1746

; CURRENT APPLICATION NUMBER: US/09/773,926

; CURRENT FILING DATE: 2001-02-02

; PRIOR APPLICATION NUMBER: 60/181,928

; PRIOR FILING DATE: 2000-02-11

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

; LENGTH: 1037

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-773-926-8

Alignment Scores:

Pred. No.: 74.9 Length: 1037

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 10.13% Indels: 0

DB: 9 Gaps: 0

US-09-972-032-2 (1-79) x US-09-773-926-8 (1-1037)

Qy 26 LeuGlyGluGlnGluGlyGly 33

Db 910 CTGGGGGAGGCGAGGAGGAGGC 887

RESULT 14

US-10-021-723A-7

; Sequence 7, Application US/10021723A

; Publication No. US20030101476A1

; GENERAL INFORMATION:  
; APPLICANT: Short, Jay

; APPLICANT: Mathur, Eric

; APPLICANT: Richardson, Toby

; APPLICANT: Robertson, Dan

; APPLICANT: Barton, Nelson

; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof

; FILE REFERENCE: 112766.140 (DIV-016Cip)

; CURRENT APPLICATION NUMBER: US/10/021,723A

; CURRENT FILING DATE: 2002-10-23

; PRIOR APPLICATION NUMBER: US 60/255,090

; PRIOR FILING DATE: 2000-12-12

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 1230

; TYPE: DNA

; ORGANISM: Rhizobium

US-10-021-723A-7

Alignment Scores:

Pred. No.: 85.4 Length: 1230

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 10.13% Indels: 0

DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-021-723A-7 (1-1230)

Qy 47 AlaSerArgTrpProArgSerAla 54  
Db 725 GCAAGCCGATGCCGAGTCGGCT 748

RESULT 15

US-10-282-122A-13867  
; Sequence 13867, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13867  
; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Burkholderia mallei  
US-10-282-122A-13867

Alignment Scores:  
Pred. No.: 85.9 Length: 1239  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-13867 (1-1239)

Qy 48 SerArgTrpProArgSerAlaSer 55  
Db 95 AGCAGATGGCGGCGAGCGCTTCA 118

Search completed: July 12, 2004, 21:27:05  
Job time : 2948 secs

\_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 12, 2004, 18:17:11 ; Search time 72 Seconds  
(without alignments)  
608.904 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 79  
Sequence: 1 MCGPRVSGCGFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 582709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1359572

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09972032/runat\_06072004\_121535\_8903/app\_query.fasta\_1.263  
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09972032@cgn 1.1.69 @runat\_06072004\_121535\_8903 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	9	11.4	1402	1	US-08-447-965A-1
2	8	10.1	1335	2	US-08-107-876-25
3	8	10.1	1335	4	US-09-295-820-25
C 4	8	10.1	1539	4	US-09-489-039A-5665
5	8	10.1	1608	4	US-09-252-991A-5081
6	8	10.1	1617	4	US-09-252-991A-5051
C 7	8	10.1	1914	4	US-09-291-922-19
8	8	10.1	2283	4	US-09-252-991A-15346
C 9	8	10.1	2463	4	US-09-252-991A-15430
10	8	10.1	2784	4	US-09-252-991A-15252
C 11	8	10.1	4403765	3	US-09-103-840A-2
C 12	8	10.1	4411529	3	US-09-103-840A-1

13	7	8.9	53	3	US-09-344-888A-3	Sequence 3, Appl1
14	7	8.9	223	6	5244792-9	Patent No. 5244792
C 15	7	8.9	262	4	US-09-280-116-53	Sequence 53, Appl
16	7	8.9	291	4	US-09-313-294A-4498	Sequence 4498, Ap
C 17	7	8.9	295	4	US-09-313-294A-4599	Sequence 5899, Ap
18	7	8.9	315	4	US-09-382-552-36	Sequence 36, Appl
C 19	7	8.9	378	4	US-09-252-991A-15117	Sequence 15117, A
20	7	8.9	392	4	US-09-621-976-14323	Sequence 14323, A
21	7	8.9	401	4	US-09-621-976-12312	Sequence 12312, A
C 22	7	8.9	432	4	US-09-252-991A-10246	Sequence 10246, A
23	7	8.9	437	4	US-09-621-976-1458	Sequence 1458, Ap
C 24	7	8.9	471	4	US-09-252-991A-14813	Sequence 14813, A
25	7	8.9	471	4	US-09-621-976-1982	Sequence 1982, Ap
26	7	8.9	476	4	US-09-621-976-15482	Sequence 15482, A
27	7	8.9	477	4	US-09-252-991A-6345	Sequence 6345, Ap
28	7	8.9	501	4	US-09-252-991A-4653	Sequence 4653, Ap
29	7	8.9	510	4	US-09-252-991A-2991	Sequence 2991, Ap
C 30	7	8.9	543	4	US-09-252-991A-16094	Sequence 16094, A
C 31	7	8.9	546	4	US-09-621-976-2743	Sequence 2743, Ap
32	7	8.9	558	4	US-09-621-976-697	Sequence 697, App
33	7	8.9	561	4	US-09-252-991A-368	Sequence 368, App
34	7	8.9	567	4	US-09-252-991A-14500	Sequence 14500, A
C 35	7	8.9	582	4	US-09-252-991A-14858	Sequence 14858, A
C 36	7	8.9	663	4	US-09-280-116-465	Sequence 265, App
37	7	8.9	678	4	US-09-252-991A-168	Sequence 168, App
38	7	8.9	699	4	US-09-252-991A-523	Sequence 523, App
39	7	8.9	765	4	US-09-489-039A-6578	Sequence 7908, Ap
C 40	7	8.9	771	4	US-09-252-991A-597	Sequence 597, App
C 41	7	8.9	789	4	US-09-252-991A-9465	Sequence 9465, Ap
C 42	7	8.9	825	4	US-09-252-991A-5418	Sequence 5418, Ap
C 43	7	8.9	834	4	US-09-252-991A-7643	Sequence 7643, Ap
C 44	7	8.9	840	4	US-09-252-991A-1634	Sequence 1634, Ap
45	7	8.9				

ALIGNMENTS

RESULT 1  
US-08-447-965A-1/c  
; Sequence 1, Application US/08447965A  
; Patent No. 5776692  
; GENERAL INFORMATION:  
; APPLICANT: El-Zaatari, Fouad A.K.  
; APPLICANT: Graham, David Y.  
; APPLICANT: Naser, Saleh  
; TITLE OF INVENTION: Mycobacterial Genus-Specific DNA Probe  
; TITLE OF INVENTION: and its Expressed Product  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fulbright & Jaworski L.L.P.  
; STREET: 801 Pennsylvania Ave. N.W.  
; CITY: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/447,965A  
; FILING DATE: 23-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davis, Peter J.  
; REGISTRATION NUMBER: 36,119  
; REFERENCE/DOCKET NUMBER: 311.019  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-662-4669  
; TELEFAX: 202-662-4643  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1402 base pairs



Percent Similarity: 100.00%    Conservatives: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 10.13%    Indels: 0  
DB: 4    Gaps: 0

US-09-972-032-2 (1-79) x US-09-295-820-25 (1-1335)

QY 51 ProArgSerAlaSerArgTrpPro 58  
DB 751 CGCGGTGCGCATCTCGATGGCG 774

## RESULT 4

US-09-489-039A-5665/c  
; Sequence 5665, Application US/09489039A  
; Patent No. 6610836

## GENERAL INFORMATION:

; APPLICANT: Gary Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 5665

; LENGTH: 1539

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-5665

Alignment Scores:  
Pred. No.: 93.5    Length: 1539  
Score: 8.00    Matches: 8  
Percent Similarity: 100.00%    Conservatives: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 10.13%    Indels: 0  
DB: 4    Gaps: 0

US-09-972-032-2 (1-79) x US-09-489-039A-5665 (1-1539)

QY 48 SerArgTrpProArgSerAlaSer 55  
DB 1162 AGCAGATGGCGCGGTGAGTTCA 1139

## RESULT 5

US-09-252-991A-5081

; Sequence 5081, Application US/09252991A

; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5081

; LENGTH: 1608

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5081

Alignment Scores:  
Pred. No.: 97.2    Length: 1608  
Score: 8.00    Matches: 8  
Percent Similarity: 100.00%    Conservatives: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 10.13%    Indels: 0

DB: 4    Gaps: 0  
US-09-972-032-2 (1-79) x US-09-252-991A-5081 (1-1608)

QY 48 SerArgTrpProArgSerAlaSer 55  
DB 668 TCGAGGTGGCGCGGTGAGTCC 691

## RESULT 6

US-09-252-991A-5051

; Sequence 5051, Application US/09252991A

; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5051

; LENGTH: 1617

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5051

Alignment Scores:  
Pred. No.: 97.6    Length: 1617  
Score: 8.00    Matches: 8  
Percent Similarity: 100.00%    Conservatives: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 10.13%    Indels: 0  
DB: 4    Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-5051 (1-1617)

QY 48 SerArgTrpProArgSerAlaSer 55  
DB 393 TCGAGGTGGCGCGGTGAGTCC 416

## RESULT 7

US-09-291-922-19/c

; Sequence 19, Application US/09291922

; Patent No. 6383776

## GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Hitz, Bill

; APPLICANT: Kinney, Tony

; APPLICANT: Tingey, Scott

; TITLE OF INVENTION: Plant Sugar Transport Proteins

; FILE REFERENCE: BB-1163

; CURRENT APPLICATION NUMBER: US/09/291,922

; CURRENT FILING DATE: 1999-04-14

; EARLIER APPLICATION NUMBER: 60/083,044

; EARLIER FILING DATE: April 24, 1998

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 19

; LENGTH: 1914

; TYPE: DNA

; ORGANISM: Zea mays

US-09-291-922-19

Alignment Scores:  
Pred. No.: 113    Length: 1914  
Score: 8.00    Matches: 8  
Percent Similarity: 100.00%    Conservatives: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 10.13%    Indels: 0

```
DB:                                     4          Gaps: 0
US-09-972-032-2 (1-79) x US-09-291-922-19 (1-1914)
QY      25 GlyLeuGlyCluGlyGlnGluGly 32
DB      1393 GGCCTTGGAGGGGACAGGAAGT 1370

RESULT 8
US-09-252-991A-15346
; Sequence 15346, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15346
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15346
Alignment Scores:
Pred. No.: 133          Length: 2283
Score: 8.00            Matches: 8
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 10.13%              Indels: 0
DB: 4                  Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-15346 (1-2283)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1218 GCTTCGAGGTGGCCCTGGAGCGCA 1241

RESULT 9
US-09-252-991A-15430/c
; Sequence 15430, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15430
; LENGTH: 2463
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15430
Alignment Scores:
Pred. No.: 142          Length: 2463
Score: 8.00            Matches: 8
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 10.13%              Indels: 0
DB: 4                  Gaps: 0

DB:                                     4          Gaps: 0
US-09-972-032-2 (1-79) x US-09-252-991A-15430 (1-2463)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1073 GCTTCGAGGTGGCCCTGGAGCGCA 1050

RESULT 10
US-09-252-991A-15252
; Sequence 15252, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15252
; LENGTH: 2784
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15252
Alignment Scores:
Pred. No.: 158          Length: 2784
Score: 8.00            Matches: 8
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 10.13%              Indels: 0
DB: 4                  Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-15252 (1-2784)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1787 GCTTCGAGGTGGCCCTGGAGCGCA 1810

RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Alignment Scores:
Pred. No.: 1.05e+05      Length: 4403765
Score: 8.00             Matches: 8
Percent Similarity: 100.00%    Conservatives: 0
```



```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x US-09-103-840A-2 (1-4403765)

QY 51 ProArgSerAlaSerArgTrpPro 58
DB 943238 CCGCGTTCGCATCGCGCTGGCGG 943215

RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: Length: 4411529
Score: 1.05e+05
Percent Similarity: 8.00%
Conservative: 0
Best Local Similarity: 100.00%
Mismatches: 0
Query Match: 10.13%
Indels: 0
Gaps: 0

US-09-972-032-2 (1-79) x US-09-103-840A-1 (1-4411529)

QY 51 ProArgSerAlaSerArgTrpPro 58
DB 943348 CCGCGTTCGCATCGCGCTGGCGG 943325

RESULT 13
US-09-344-888A-3
; Sequence 3, Application US/09344888A
; Patent No. 6291245
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Ethard
; APPLICANT: Schantz, Christian
; TITLE OF INVENTION: New Host-Vector System
; FILE REFERENCE: CD20315
; CURRENT APPLICATION NUMBER: US/09/344,888A
; CURRENT FILING DATE: 1999-06-25
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 53
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:primer
US-09-344-888A-3

Alignment Scores:
Pred. No.: Length:
Score: 165
7.00

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x US-09-344-888A-3 (1-53)

QY 27 GlyGluGlyGlnGluGlyGly 33
DB 32 CGAGAGGCCCAAGAGGAGGG 52

RESULT 14
US-09-972-032-2 (1-79) x US-09-344-888A-3 (1-53)

QY 27 GlyGluGlyGlnGluGlyGly 33
DB 32 CGAGAGGCCCAAGAGGAGGG 52

RESULT 15
US-09-280-116-53/c
; Sequence 53, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: aspartyl proteases
; NAME/KEY: misc feature
; LOCATION: (1)-(262)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-53

Alignment Scores:
Pred. No.: Length:
Score: 165
7.00

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.86% Indels: 0
DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x 5244792-9 (1-223)

QY 50 TrpProArgSerAlaSerArg 56
DB 202 TGGCGCGGTCGGCGTCGCGC 222

RESULT 15
US-09-280-116-53/c
; Sequence 53, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: aspartyl proteases
; NAME/KEY: misc feature
; LOCATION: (1)-(262)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-53

Alignment Scores:
Pred. No.: Length:
Score: 165
7.00
```

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.86% Indels: 0  
 DB: 4 Gaps: 0

US-09-972-032-2 (1-79) x US-09-280-116-53 (1-262)

Qy 69 ProGlnLeuGlyGluLeuCys 75  
 ||||||||||||||||  
 Db 169 CCCCACTTGGGGAGCTCTGC 149

Search completed: July 12, 2004, 18:46:21  
 Job time : 1734 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 11:56:49 ; Search time 495 Seconds  
(without alignments)  
8496.396 Million cell updates/sec

Title: US-09-972-032-1  
Perfect score: 990  
Sequence: 1 ggaatgttctcgaggccaa.....aaaatgagcggcgcaagt 990

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1990s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	6	ABL60606 Human ERC
2	147	14.8	597	5	AAS69040 DNA encod
3	80	8.1	404	4	AAL01097 Human rep
4	80	8.1	404	4	ABL96563 Human tes
5	51	5.2	149480	6	ABL61947 Colon ade
6	51	5.2	149480	6	ABL68365 Kidney ca
7	51	5.2	149480	6	ABL61948 Colon ade
8	48	4.8	118584	9	ADC87623 Human GPC
9	48	4.8	349999	9	ADC87010 Human GPC
10	47	4.7	494	5	ABA15806 Human ner
11	47	4.7	9888	4	AAK71311 Human imm
12	47	4.7	13160	5	ABA15804 Human ner
13	47	4.7	13189	5	ABA15805 Human ner
14	47	4.7	43419	9	ADC86998 Human GPC
15	47	4.7	63626	8	ADC27000 Human car
16	47	4.7	63720	8	ADA03080 Human PRD
17	47	4.7	63720	8	ADA66364 Human PRD
18	47	4.7	63720	8	ADB72818 Human PRD
19	47	4.7	103747	6	ABQ86139 Human osc
20	47	4.7	122748	6	ABT10719 Human bre
21	46	4.6	274	3	AAC19185 Human sec
22	46	4.6	394	4	AAI82243 Human pol
23	46	4.6	2105	4	AAH99495 Human pro

24	46	4.6	2904	9	ADB63202	Adb63202 Human CDN
25	46	4.6	2940	8	AAD58740	Aad58740 Human tra
c	26	4.6	8865	4	AAK87029	Aak87029 Human imm
27	46	4.6	12117	3	AA96368	Aa96368 Polymorph
28	46	4.6	22452	4	AAS27829	Aas27829 DNA encod
29	46	4.6	22452	4	AAS27827	Aas27827 DNA encod
30	46	4.6	22452	9	ADB94632	Adb94632 Novel hum
31	46	4.6	22452	9	ADB94630	Adb94630 Novel hum
32	46	4.6	32193	4	AAD16595	Aad16595 Human nov
33	46	4.6	32193	4	AAI36258	Aai36258 Human mus
34	46	4.6	32193	7	ABX59246	Abx59246 cDNA enco
35	46	4.6	32249	4	AAL04676	Aal04676 Human rep
36	46	4.6	32249	4	ABL97583	Abi97583 Human tes
c	37	46	50000	3	AA96363	Aa96363 Polymorph
38	46	4.6	75270	7	ACA63030	Ac63030 Human chr
39	46	4.6	75270	9	ADD71050	Add71050 Human sec
40	45	4.5	122	4	AAK84551	Aak84551 Human imm
c	41	45	308	4	AAS32644	Aas32644 Human gen
c	42	45	354	4	AAS37387	Aas37387 Novel hum
c	43	45	414	4	AAI89811	Aai89811 Human pol
c	44	45	444	5	ABV53291	Abv53291 Human pro
45	45	4.5	482	4	AAI91309	Aai91309 Human pol

## ALIGNMENTS

## RESULT 1

ABL60606  
ID ABL60606 standard; cDNA; 990 BP.  
XX  
AC ABL60606;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human ERCoA3 protein encoding cDNA.  
XX  
KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;  
KW osteoporosis; cytostatic; osteopathic; human; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 203..442  
FT /\*tag= a  
FT /product= "ERCoA3 protein"  
XX  
XX WO200228352-A2.  
XX 11-APR-2002.  
XX 05-OCT-2001; 2001WO-US031271.  
XX 05-OCT-2000; 2000US-0238190P.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX Montano M, Sutton A;  
XX WPI: 2002-454492/48.  
XX P-PSDB; ABB08035.  
XX  
XX New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen  
XX Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or  
XX estrogen-induced proliferation of cancer cells and in treating  
XX osteoporosis.  
XX Claim 5; Fig 1; 39pp; English.  
XX  
XX The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3)  
XX protein and encoding polynucleotides. ERCoA3 can be used to inhibit or  
XX reduce tamoxifen or estrogen induced proliferation of cancer cells, by  
XX reduced activity of ERCoA3, and for detecting cancer cells that are

CC tamoxifen resistant, or to treat osteoporosis, by increasing levels of  
CC ERCoA3 in cells. The encoding polynucleotide can be used to inhibit  
CC translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator  
CC protein and can bind to the estrogen receptor to activate a molecular  
CC cellular response in the cell. The present sequence represents the human  
CC ERCoA3 encoding cDNA  
XX  
SQ Sequence 990 BP; 189 A; 281 C; 317 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 990; DB 6; Length 990;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAATTGTTCTCGAGGCCAAGAATTGGCAAGCGCATCTCTGACCTCGTGATCCGCCCG 60  
DB 1 GGAATTGTTCTCGAGGCCAAGAATTGGCAAGCGCATCTCTGACCTCGTGATCCGCCCG 60

QY 61 CTTCAGGCTCCCAAGTCTGGGATACAGGCATGAGCCACCGCGCCCGCCCGACAC 120  
DB 61 CTTCAGGCTCCCAAGTCTGGGATACAGGCATGAGCCACCGCGCCCGCCCGACAC 120

QY 121 CTAGTTTAAAGGCCCTCTGTTGCTGCGCTGCGCGCTCCCGAGCTGCCAGTCTG 180  
DB 121 CTAGTTTAAAGGCCCTCTGTTGCTGCGCTGCGCGCTCCCGAGCTGCCAGTCTG 180

QY 181 GCGGGCTCAGTCCCGGTTGCCATGTGTGGAGACCGCGTTCGCTGAAGCGCTGGATGG 240  
DB 181 GCGGGCTCAGTCCCGGTTGCCATGTGTGGAGACCGCGTTCGCTGAAGCGCTGGATGG 240

QY 241 CTTCGCTGATGCATTTGACCGGCTCTGGACTGGCTAGGAGAGGACGAGGGCGG 300  
DB 241 CTTCGCTGATGCATTTGACCGGCTCTGGACTGGCTAGGAGAGGACGAGGGCGG 300

QY 301 AATTGGCCCCAGGGCCAGTCTCGCCGACCCCGACTCGCGCTCCCGTGGCCCGCGAG 360  
DB 301 AATTGGCCCCAGGGCCAGTCTCGCCGACCCCGACTCGCGCTCCCGTGGCCCGCGAG 360

QY 361 CGCTCCCGTGGCTGGAGTGCAGTCTTACCGTCCGAGATCGTCCGCACTGGGCA 420  
DB 361 CGCTCCCGTGGCTGGAGTGCAGTCTTACCGTCCGAGATCGTCCGCACTGGGCA 420

QY 421 GCTGTGATCGGGCGTGGCTTAAGCCGCTGTTGGTTAGATTGGCCAGCGGACTTAAG 480  
DB 421 GCTGTGATCGGGCGTGGCTTAAGCCGCTGTTGGTTAGATTGGCCAGCGGACTTAAG 480

QY 481 TGTGTCTCTGAAGAGATGGAATTAAGTCTGAGGGTCTCTGAAGAGTATCCCGCC 540  
DB 481 TGTGTCTCTGAAGAGATGGAATTAAGTCTGAGGGTCTCTGAAGAGTATCCCGCC 540

QY 541 CACATCAATGGCTTAGTCTAGGAGCGGCTGGTGGGGCTTAGGGCGAGGGG 600  
DB 541 CACATCAATGGCTTAGTCTAGGAGCGGCTGGTGGGGCTTAGGGCGAGGGG 600

QY 601 CAGACATACCCCAAGTGGTGAATTGTATACCGCAAGGGCTGGATCGAACCCGCCAA 660  
DB 601 CAGACATACCCCAAGTGGTGAATTGTATACCGCAAGGGCTGGATCGAACCCGCCAA 660

QY 661 GACATCGAAGGCTGTGCTGCTGAGAGGCGCGCAATCCAGTGTGCTGGGCTTAC 720  
DB 661 GACATCGAAGGCTGTGCTGCTGAGAGGCGCGCAATCCAGTGTGCTGGGCTTAC 720

QY 721 AGAAAGAGCTCCACCTTCTTGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780  
DB 721 AGAAAGAGCTCCACCTTCTTGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780

QY 781 GAGCTGGGGCGGGGAGATGCTGCCCGCAGTACAAAGCTGATTGACCTGGGGCTCTG 840  
DB 781 GAGCTGGGGCGGGGAGATGCTGCCCGCAGTACAAAGCTGATTGACCTGGGGCTCTG 840

QY 841 GACTTCCCTGATTCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCGCTGCCTTCAT 900  
DB 841 GACTTCCCTGATTCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCGCTGCCTTCAT 900

## RESULT 2

AAS69040/c

ID AAS69040 standard; cDNA; 597 BP.

XX

AC AAS69040;

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #4844.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

XX

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR F-PSDB; AEG04853.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

XX

PT diagnostics, forensics, gene mapping, identification of mutations

XX

PT responsible for genetic disorders or other traits and to assess

XX

PT biodiversity.

XX

PS Claim 1; SEQ ID NO 4844; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX

CC sequences. (I) is useful as hybridisation probes, polymerase chain

XX

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX

CC and in recombinant production of (II). The polynucleotides are also used

XX

CC in diagnostics as expressed sequence tags for identifying expressed

XX

CC genes. (I) is useful in gene therapy techniques to restore normal

XX

CC activity of (II) or to treat disease states involving (II). (II) is

XX

CC useful for generating antibodies against it, detecting or quantitating a

XX

CC polypeptide in tissue, as molecular weight markers and as a food

XX

CC supplement. (II) and its binding partners are useful in medical imaging

XX

CC of sites expressing (II). (I) and (II) are useful for treating disorders

XX

CC involving aberrant protein expression or biological activities. The

XX

CC polypeptide and polynucleotide sequences have applications in

XX

CC diagnostics, forensics, gene mapping, identification of mutations

XX

CC responsible for genetic disorders or other traits to assess biodiversity

XX

CC and to produce other types of data and products dependent on DNA and

XX

CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic

XX

CC coding sequences of the invention. Note: The sequence data for this

XX

CC patent did not appear in the printed specification, but was obtained in

XX

CC electronic format directly from WIPO at

XX

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;

Query Match 14.8%; Score 147; DB 5; Length 597;

Best Local Similarity 100.0%; Pred. No. 4e-48;



PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249220P.  
PR 17-NOV-2000; 2000US-0249221P.  
PR 17-NOV-2000; 2000US-0249222P.  
PR 17-NOV-2000; 2000US-0249223P.  
PR 17-NOV-2000; 2000US-0249224P.  
PR 17-NOV-2000; 2000US-0249225P.  
PR 17-NOV-2000; 2000US-0249226P.  
PR 17-NOV-2000; 2000US-0249227P.  
PR 17-NOV-2000; 2000US-0249228P.  
PR 17-NOV-2000; 2000US-0249229P.  
PR 17-NOV-2000; 2000US-0249230P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251388P.  
PR 05-DEC-2000; 2000US-0251619P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX P-PSDB; AAM95127.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX Claim 1; SEQ ID NO 1098; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders.  
XX including cancer. The present sequence is a coding sequence of the  
XX invention  
XX  
XX Sequence 404 BP; 83 A; 109 C; 107 G; 95 T; 0 U; 10 Other;  
SQ  
Query Match 8.1%; Score 80; DB 4; Length 404;  
Best Local Similarity 100.0%; Pred. No. 8.3e-22;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 872 CAAAGTCCTGCGGTGGTGGCTTCACCACTCTCTCACTTCTGCTTCAGAGTAA 931  
Db 8 CAAAGTCCTGCGGTGGTGGCTTCACCACTCTCTCACTTCTGCTTCAGAGTAA 67  
Qy 932 AATTGCAAGATCTGTGTGC 951  
Db 68 AATTGCAAGATCTGTGTGC 87  
RESULT 4  
ABL96563  
ID ABL96563 standard; cDNA; 404 BP.  
XX  
XX ABL96563;  
XX  
XX 21-JUN-2002 (first entry)  
XX  
XX Human testicular antigen encoding cDNA SEQ ID NO: 231.  
XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
XX reproductive system disorder; urinary system disorder; gene therapy;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
KW  
KW

KW gastrointestinal disease; infection; cytostatic; gene; ss.  
XX Homo sapiens.  
CS WO200155317-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001329.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 14-AUG-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225256P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225477P.  
XX 14-AUG-2000; 2000US-0225577P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
XX 14-SEP-2000; 2000US-0232401P.  
XX 14-SEP-2000; 2000US-0233063P.  
XX 14-SEP-2000; 2000US-0233064P.  
XX 14-SEP-2000; 2000US-0233065P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234997P.



PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 03-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 DR WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1; SEQ ID NO 284; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABU61664  
 CC to ABU70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'  
 CC tumour  
 XX  
 SQ Sequence 149480 BP; 38770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;  
 Query Match 5.2%; Score 51; DB 6; Length 149480;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-11;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 57 CCGCCTCAGCTTCCCAAGTGTGGATTACAGGATGAGCCACCGGCC 107  
 DB 76386 CCGCCTCAGCTTCCCAAGTGTGGATTACAGGATGAGCCACCGGCC 76436  
 RESULT 6  
 ABU68365  
 ID ABU68365 standard; DNA; 149480 BP.

XX ABL68365;  
 AC 15-MAY-2002 (first entry)  
 XX  
 DT Kidney cancer related gene sequence SEQ ID NO:6702.  
 XX  
 DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 DR WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.



XX Claim 1; SEQ ID NO 6702; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour  
XX  
SQ Sequence 149480 BP; 39770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;  
Query Match 5.2%; Score 51; DB 6; Length 149480;  
Best Local Similarity 100.0%; Pred.No. 7.6e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 57 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGGCC 107  
Db 76386 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGGCC 76436  
RESULT 7  
ABL61948  
ID ABL61948 standard; DNA; 149480 BP.  
XX  
AC ABL61948;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Colon adenocarcinoma related gene sequence SEQ ID NO:285.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200194629-A2.  
XX  
XX 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US010838.  
XX  
PR 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0231133P.  
PR 18-SEP-2000; 2000US-0231617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 29-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.  
XX  
XX Claim 1; SEQ ID NO 285; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour  
XX  
SQ Sequence 149480 BP; 39770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;  
XX  
XX Query Match 5.2%; Score 51; DB 6; Length 149480;  
XX Best Local Similarity 100.0%; Pred.No. 7.6e-11;  
XX Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 57 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGGCC 107  
Db 76386 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGGCC 76436  
RESULT 8  
ADC87623/c  
ID ADC87623 standard; DNA; 118584 BP.  
XX  
AC ADC87623;  
XX

DT 01-JAN-2004 (first entry)  
 DE Human GPCR related polynucleotide SEQ ID NO:2076.  
 XX ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;  
 KW gene therapy.  
 XX Homo sapiens.  
 OS  
 XX EP1270724-A2.  
 FN  
 XX 02-JAN-2003.  
 PD  
 XX  
 XX 18-JUN-2002; 2002EP-00013517.  
 PF  
 XX 18-JUN-2001; 2001JP-00246789.  
 PR  
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
 PA Suwa M, Asai K, Akiyama Y, Aburatani H;  
 PI WPI; 2003-315783/31.  
 DR  
 XX New polynucleotide, useful for preparing a composition for treating a  
 XX patient in need of increased or suppressed activity or expression of the  
 XX guanosine triphosphate-binding protein coupled receptor.  
 XX Disclosure; SEQ ID NO 2076; 28pp; English.  
 XX The invention relates to a novel polynucleotide encoding a guanosine  
 XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
 XX the invention may have a use in gene therapy. The polynucleotide and  
 XX polypeptide are useful for preparing a composition for treating a patient  
 XX in need of increased or suppressed activity or expression of the  
 XX guanosine triphosphate-binding protein coupled receptor. The protein  
 XX sequences shown in ADC87618-ADC87623 represent polynucleotide sequences  
 XX related to the invention.  
 XX Sequence 118584 BP; 32219 A; 27585 C; 27306 G; 31474 T; 0 U; 0 Other;  
 SQ  
 Query Match 4.8%; Score 48; DB 9; Length 118584;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX 57 CCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104  
 QY 29455 CCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 29408  
 DB  
 RESULT 9  
 ADC87010/c  
 ID ADC87010 standard; DNA; 349999 BP.  
 XX  
 AC ADC87010;  
 XX  
 DT 01-JAN-2004 (first entry)  
 DE Human GPCR gene SEQ ID NO:1463.  
 XX  
 DE ds; gene; human; GPCR;  
 KW guanosine triphosphate-binding protein coupled receptor; gene therapy.  
 KW Homo sapiens.  
 OS  
 XX EP1270724-A2.  
 FN  
 XX 02-JAN-2003.  
 PD  
 XX 18-JUN-2002; 2002EP-00013517.  
 PF  
 XX 19-JUN-2001; 2001JP-00246789.  
 PR  
 XX

PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
 PI WPI; 2003-315783/31.  
 DR P-PSDB; ADC87011.  
 DR  
 XX New polynucleotide, useful for preparing a composition for treating a  
 XX patient in need of increased or suppressed activity or expression of the  
 XX guanosine triphosphate-binding protein coupled receptor.  
 XX Claim 1; SEQ ID NO 1463; 28pp; English.  
 XX The invention relates to a novel polynucleotide encoding a guanosine  
 XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
 XX the invention may have a use in gene therapy. The polynucleotide and  
 XX polypeptide are useful for preparing a composition for treating a patient  
 XX in need of increased or suppressed activity or expression of the  
 XX guanosine triphosphate-binding protein coupled receptor. The  
 XX polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
 XX invention.  
 XX Sequence 349999 BP; 91105 A; 78863 C; 81174 G; 98857 T; 0 U; 0 Other;  
 SQ  
 Query Match 4.8%; Score 48; DB 9; Length 349999;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-10;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 57 CCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104  
 DB 329421 CCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 329374  
 RESULT 10  
 ABA15806  
 ID ABA15806 standard; DNA; 494 BP.  
 XX  
 AC ABA15806;  
 XX  
 DT 23-JAN-2002 (first entry)  
 DE Human nervous system related polynucleotide SEQ ID NO 8137.  
 XX  
 KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;  
 KW antineumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 XX Homo sapiens.  
 OS  
 XX WO200159063-A2.  
 FN  
 XX 16-AUG-2001.  
 PD  
 XX  
 PF 17-JAN-2001; 2001WO-US001334.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.

PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226686P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229387P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231242P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232297P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0233063P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234937P.
PR	25-SEP-2000;	2000US-0234938P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	29-SEP-2000;	2000US-0235836P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	20-OCT-2000;	2000US-0242221P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249219P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250391P.
PR	01-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-541565/60.	
XX		
DR		
XX		
XX		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system cancers	
PT	and metastases.	
XX		
PS	Disclosure; SEQ ID NO 8137; 1701pp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (AB211004-AB21534) and proteins	
CC	(AB214678-AB218001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in	
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,	
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 494 BP; 99 A; 124 C; 121 G; 150 T; 0 U; 0 Other;

Query Match 4.7%; Score 47; DB 5; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAGTCTGGATTACAGCATGAGCCACCGCGCGGCCCCC 115  
 Db 259 TCCCAAGTCTGGATTACAGCATGAGCCACCGCGCGGCCCCC 305

## RESULT 11

AAK71311  
 ID AAK71311 standard; DNA; 9888 BP.

XX AAK71311;

XX 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26123.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW Cytostatic; Gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US0001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184654P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214986P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

XX 22-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226868P.

PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228944P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234957P.  
 PR 25-SEP-2000; 2000US-0234988P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 01-NOV-2000; 2000US-0244618P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 17-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.

PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-483426/52.  
DR  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
PT  
XX  
XX  
PS Disclosure; SEQ ID NO 26123; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX  
SQ Sequence 9888 BP; 2845 A; 1902 C; 2021 G; 3120 T; 0 U; 0 Other;  
  
Query Match 4.7%; Score 47; DB 4; Length 9888;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 69 TCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCGCGCGCC 115  
|||||  
Db 5540 TCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCGCGCC 5586  
  
RESULT 12  
ABAI5804

ID  
XX  
XX ABAI5804 standard; DNA; 13160 BP.  
AC ABAI5804;  
XX  
DT 23-JAN-2002 (first entry)  
DE  
XX Human nervous system related polynucleotide SEQ ID NO 8135.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischizoid; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225265P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226661P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.

```
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 21-SEP-2000; 2000US-0234977P.
PR 21-SEP-2000; 2000US-0234998P.
PR 21-SEP-2000; 2000US-0234999P.
PR 21-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239235P.
PR 13-OCT-2000; 2000US-0239236P.
PR 13-OCT-2000; 2000US-0239237P.
PR 20-OCT-2000; 2000US-0240936P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system cancers
XX and metastases.
XX
XX Disclosure; SEQ ID NO 8135; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (ABB14678-ABB18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 13160 BP; 3247 A; 2880 C; 3120 G; 3913 T; 0 U; 0 Other;

Query Match 4.7%; Score 47; DB 5; Length 13160;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCCCAAAGTGTGGATTACAGCATGAGCCAGCCGCGCGCGCCCC 115
|||
Db 8089 TCCCAAAGTGTGGATTACAGCATGAGCCAGCCGCGCGCCCC 8135

RESULT 13
ABA15805
ID ABA15805 standard; DNA; 13189 BP.
XX
XX ABA15805;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 8136.
XX
XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
XX antiparkinsonian; antischlicking; antiamebic; antithratic; cancer;
XX antineumatic; hepatotropic; cerebroprotective; antinflammatory;
XX antiallergic; antidiabetic; antituler; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
```

XX	Homo sapiens.	26-SEP-2000;	2000US-0235484P	XX
OS		27-SEP-2000;	2000US-0235834P	PR
XX		27-SEP-2000;	2000US-0235836P	PR
XX		29-SEP-2000;	2000US-0236327P	PR
PN	WO200159063-A2.	29-SEP-2000;	2000US-0236367P	PR
XX		29-SEP-2000;	2000US-0236368P	PR
PD	16-AUG-2001.	29-SEP-2000;	2000US-0236369P	PR
XX		29-SEP-2000;	2000US-0236370P	PR
PF	17-JAN-2001; 2001WO-US0001334.	02-OCT-2000;	2000US-0236802P	PR
XX		02-OCT-2000;	2000US-0237037P	PR
XX		02-OCT-2000;	2000US-0237038P	PR
PR		02-OCT-2000;	2000US-0237039P	PR
PR		02-OCT-2000;	2000US-0237040P	PR
PR		13-OCT-2000;	2000US-0239335P	PR
PR		13-OCT-2000;	2000US-0239337P	PR
PR		20-OCT-2000;	2000US-0240960P	PR
PR		20-OCT-2000;	2000US-0241785P	PR
PR		20-OCT-2000;	2000US-0241786P	PR
PR		20-OCT-2000;	2000US-0241787P	PR
PR		20-OCT-2000;	2000US-0241808P	PR
PR		20-OCT-2000;	2000US-0241809P	PR
PR		20-OCT-2000;	2000US-0241826P	PR
PR		20-OCT-2000;	2000US-0242221P	PR
PR		01-NOV-2000;	2000US-0244617P	PR
PR		08-NOV-2000;	2000US-0246474P	PR
PR		08-NOV-2000;	2000US-0246475P	PR
PR		08-NOV-2000;	2000US-0246476P	PR
PR		08-NOV-2000;	2000US-0246477P	PR
PR		08-NOV-2000;	2000US-0246478P	PR
PR		08-NOV-2000;	2000US-0246523P	PR
PR		08-NOV-2000;	2000US-0246524P	PR
PR		08-NOV-2000;	2000US-0246525P	PR
PR		08-NOV-2000;	2000US-0246526P	PR
PR		08-NOV-2000;	2000US-0246527P	PR
PR		08-NOV-2000;	2000US-0246528P	PR
PR		08-NOV-2000;	2000US-0246532P	PR
PR		08-NOV-2000;	2000US-0246609P	PR
PR		08-NOV-2000;	2000US-0246610P	PR
PR		08-NOV-2000;	2000US-0246611P	PR
PR		08-NOV-2000;	2000US-0246613P	PR
PR		17-NOV-2000;	2000US-0249207P	PR
PR		17-NOV-2000;	2000US-0249208P	PR
PR		17-NOV-2000;	2000US-0249209P	PR
PR		17-NOV-2000;	2000US-0249210P	PR
PR		17-NOV-2000;	2000US-0249211P	PR
PR		17-NOV-2000;	2000US-0249212P	PR
PR		17-NOV-2000;	2000US-0249213P	PR
PR		17-NOV-2000;	2000US-0249214P	PR
PR		17-NOV-2000;	2000US-0249215P	PR
PR		17-NOV-2000;	2000US-0249216P	PR
PR		17-NOV-2000;	2000US-0249217P	PR
PR		17-NOV-2000;	2000US-0249218P	PR
PR		17-NOV-2000;	2000US-0249244P	PR
PR		17-NOV-2000;	2000US-0249245P	PR
PR		17-NOV-2000;	2000US-0249264P	PR
PR		17-NOV-2000;	2000US-0249265P	PR
PR		17-NOV-2000;	2000US-0249297P	PR
PR		17-NOV-2000;	2000US-0249299P	PR
PR		17-NOV-2000;	2000US-0249300P	PR
PR		01-DEC-2000;	2000US-0250391P	PR
PR		01-DEC-2000;	2000US-0251160P	PR
PR		05-DEC-2000;	2000US-0251030P	PR
PR		05-DEC-2000;	2000US-0251988P	PR
PR		05-DEC-2000;	2000US-0256719P	PR
PR		06-DEC-2000;	2000US-0251479P	PR
PR		08-DEC-2000;	2000US-0251856P	PR
PR		08-DEC-2000;	2000US-0251868P	PR
PR		08-DEC-2000;	2000US-0251869P	PR
PR		08-DEC-2000;	2000US-0251989P	PR
PR		11-DEC-2000;	2000US-0251930P	PR
PR		11-DEC-2000;	2000US-0254097P	PR
PR		05-JAN-2001;	2001US-0259678P	XX
XX				XX

PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX Disclosure; SEQ ID NO 8136; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins  
CC (ABAI1478-ABAI18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 13189 BP; 3249 A; 2881 C; 3117 G; 3942 T; 0 U; 0 Other;  
SQ  
Query Match 4.7%; Score 47; DB 5; Length 13189;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 TCCCAAGTCTGGGATTACAGCAGTACGACCGCGCGCGCGCC 115  
DB 8116 TCCCAAGTCTGGGATTACAGCAGTACGACCGCGCGCGCC 8162  
RESULT 14  
ADC86998/c  
ID ADC86998 standard; DNA; 43419 BP.  
XX  
XX  
AC ADC86998;  
XX  
XX 01-JAN-2004 (first entry)  
XX  
XX Human GPCR gene SEQ ID NO:1451.  
DE  
XX  
XX ds; gene; human; GPCR;  
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.  
KW  
XX  
XX Homo sapiens.  
XX  
XX EP1270724-A2.  
XX  
XX 02-JAN-2003.  
XX  
XX 18-JUN-2002; 2002BP-00013517.  
XX  
XX 18-JUN-2001; 2001JP-00245789.  
XX  
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
PI WPI; 2003-315783/31.  
XX  
XX P-PSDB; ADC86999.  
XX  
XX New polynucleotide, useful for preparing a composition for treating a

PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
XX Claim 1; SEQ ID NO 1451; 28pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.  
XX  
XX Sequence 43419 BP; 11054 A; 10220 C; 10281 G; 11764 T; 0 U; 100 Other;  
SQ  
Query Match 4.7%; Score 47; DB 9; Length 43419;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 52 ATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGCAGTACG 98  
DB 31644 ATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGCAGTACG 31598  
RESULT 15  
ADC27000  
ID ADC27000 standard; DNA; 63626 BP.  
XX  
XX  
AC ADC27000;  
XX  
XX 18-DEC-2003 (first entry)  
XX  
XX Human carcinoma associated (CA) genomic DNA PRDM11.  
XX  
XX Carcinoma associated gene; CA gene; PR domain protein 11; PRDM11;  
KW carcinoma associated protein; CAP; carcinoma; cancer; adenocarcinoma;  
KW breast cancer; inflammatory carcinoma; Paget's disease; aging;  
KW neurodegenerative disease; cytostatic; osteopathic; neurotropic;  
KW neuroprotective; human; ds.  
XX  
XX Homo sapiens.  
XX  
XX US2003087252-A1.  
XX  
XX 08-MAY-2003.  
XX  
XX 20-MAR-2002; 2002US-00105637.  
XX  
XX 22-DEC-2000; 2000US-00747377.  
XX  
XX 02-MAR-2001; 2001US-00798586.  
XX  
XX 20-DEC-2001; 2001US-00034650.  
XX  
XX (MORR/) MORRIS D W.  
FA (ENGE/) ENGELHARD E K.  
XX  
XX Morris DW, Engelhard EK;  
PI  
XX  
XX WPI; 2003-730063/69.  
XX  
XX Novel recombinant carcinoma-associated gene, PRDM11 useful for diagnosing  
PT cancer, and for screening for agents for treating cancers and  
PT neurodegenerative diseases.  
XX  
XX Claim 1; Page 36-45; 49pp; English.  
XX  
XX The present invention relates to novel carcinoma associated (CA) nucleic  
CC acid sequences from both mouse and human. In particular the invention  
CC discloses a recombinant nucleic acid comprising a fully defined PR domain  
CC protein 11 (PRDM11 - a carcinoma-associated gene) nucleotide sequence.  
CC The sequences of the invention are useful for identifying compounds that  
CC modulate the activity of a carcinoma associated protein (CAP) which are  
CC potentially useful in treating carcinomas. Among the diseases and



CC conditions that may be diagnosed or treated according to the invention  
CC are cancers such as adenocarcinoma, breast cancer, inflammatory  
CC carcinoma, Paget's disease, etc, aging, and neurodegenerative diseases.  
CC The present sequence represents a CA nucleic acid sequence of the  
CC invention.

XX

SQ Sequence 63626 BP; 15622 A; 15444 C; 15753 G; 16795 T; 0 U; 12 Other;

Query Match 4.7%; Score 47; DB 9; Length 63626;

Best Local Similarity 100.0%; Pred No. 3.2e-09;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 GCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGCGCC 112

|||||

Db 18670 GCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGCGCC 18716

Search completed: July 10, 2004, 13:27:05

Job time : 501 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 13:04:20 ; Search time 3110 Seconds  
(without alignments)  
9505.967 Million cell updates/sec

Title: US-09-972-032-1  
Perfect score: 990  
Sequence: 1 ggaattgtctcgaggccaa.....aaaatgagcgccgaagt 990

Scoring table: OLIGO\_NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	56.8	1467	11	BC039500 Homo sapi
C 2	370	37.4	760	12	BM980506
C 3	343	34.6	691	13	BU620427
C 4	341	34.4	1201	13	EX406138

C 5	323	32.6	669	12	BM679577
C 6	323	32.6	694	13	BU684594
C 7	312	31.5	623	14	CD742581
C 8	308	31.1	632	13	BU617331
C 9	276	27.9	628	13	BQ447041
C 10	269	27.2	633	14	CA313142
C 11	225	22.7	395	9	AA687318
C 12	212	21.4	300	12	BM672504
C 13	211	21.3	462	12	BM687386
C 14	211	21.3	480	10	BF509528
C 15	210	21.2	367	10	BF002915
C 16	210	21.2	411	9	AI184808
C 17	209	21.1	211	9	AI202243
C 18	207	20.9	439	10	BE677246
C 19	184	18.6	1084	12	BM806108
C 20	103	10.4	282	10	BF364571
C 21	57	5.8	751	29	AG160016
C 22	57	5.8	751	29	AG037002
C 23	55	5.6	688	29	AG118999
C 24	54	5.5	658	29	AG161334
C 25	54	5.5	740	28	AZ520059
C 26	53	5.4	172	10	AW841136
C 27	53	5.4	224	10	AW841118
C 28	53	5.4	450	28	AQ331706
C 29	48	4.8	433	10	BE243207
C 30	48	4.8	715	29	AG153081
C 31	47	4.7	328	9	AI370057
C 32	47	4.7	340	14	T02817
C 33	47	4.7	379	9	AI370094
C 34	47	4.7	379	10	AW403177
C 35	47	4.7	383	9	AI370074
C 36	47	4.7	578	14	CA397714
C 37	47	4.7	680	29	AG124506
C 38	46	4.6	155	13	BU571107
C 39	46	4.6	386	9	AS514737
C 40	46	4.6	407	10	BE152030
C 41	46	4.6	414	28	B66825
C 42	46	4.6	423	14	H60216
C 43	46	4.6	493	14	N92133
C 44	46	4.6	500	10	BF820071
C 45	46	4.6	513	28	AQ235733

#### ALIGNMENTS

LOCUS	BC039500	Homo sapiens	clone IMAGE:5555626	1467 bp	mRNA	linear	HTC 04-MAR-2003
DEFINITION	BC039500	Homo sapiens	clone IMAGE:5555626	1467 bp	mRNA	linear	HTC 04-MAR-2003
ACCESSION	BC039500	Homo sapiens	clone IMAGE:5555626	1467 bp	mRNA	linear	HTC 04-MAR-2003
VERSION	BC039500.1	GI:25058499					
KEYWORDS	HTC						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
REFERENCE	1 (bases 1 to 1467)						
AUTHORS	Strausberg, R.						
TITLE	Direct Submission						
JOURNAL	Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA						
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>						
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs@mail.nih.gov">cgapbs@mail.nih.gov</a> Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing: The I.M.A.G.E. Consortium (LLNL) Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www-sbhc.stanford.edu">http://www-sbhc.stanford.edu</a>						

Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 88 Row: 9 Column: 1  
This clone has the following problem: retained intron.

## FEATURES

source

```
1. 1467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5555626"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH MGC 71"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
```

## ORIGIN

Query Match 56.8%; Score 562; DB 11; Length 1467;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 662; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 34 CGATCTCTCGCTCGGTCGTCGCGCCGCTCAGCTTCCCAAGTGTGGATTACAGGCA 93
Db 504 CGATCTCTCGCTCGGTCGTCGCGCCGCTCAGCTTCCCAAGTGTGGATTACAGGCA 563
Qy 94 TGAGCCACCGCGCCGCGCCGCGACACCTAGTATTTAAAGGGCCCTGCTGCTGCCGC 153
Db 564 TGAGCCACCGCGCCGCGCCGCGACACCTAGTATTTAAAGGGCCCTGCTGCTGCCGC 623
Qy 154 TGGCGCGCTCCAGCTGCGCCAGTCTGGCGGCTCAGTCCCGGTGCGCATGTGGGAG 213
Db 624 TGGCGCGCTCCAGCTGCGCCAGTCTGGCGGCTCAGTCCCGGTGCGCATGTGGGAG 683
Qy 214 ACCGCGCTCGCTAAGCGCTGATGTGGCTTCGCTGATGACATTCGACCGGCTCTGAC 273
Db 684 ACCGCGCTCGCTAAGCGCTGATGTGGCTTCGCTGATGACATTCGACCGGCTCTGAC 743
Qy 274 TGGGCTAGGGGAAGGGCAGAGGGCGGAATGGGCCGAGGCGCAAGCTCCGCCACCC 333
Db 744 TGGGCTAGGGGAAGGGCAGAGGGCGGAATGGGCCGAGGCGCAAGCTCCGCCACCC 803
Qy 334 CGACTGCGCTCCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGGTCCTTAC 393
Db 804 CGACTGCGCTCCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGGTCCTTAC 863
Qy 394 CGTCCGAGATCGTCCGCAACTGGCGAGTGTGCATGGGGCGTGGCTAAGGCGCGTGT 453
Db 864 CGTCCGAGATCGTCCGCAACTGGCGAGTGTGCATGGGGCGTGGCTAAGGCGCGTGT 923
Qy 454 GGTACGATTGGCAGGGGACTTAAAGTGTGTCTCTGAAGCATGACATTAAGTCTGG 513
Db 924 GGTACGATTGGCAGGGGACTTAAAGTGTGTCTCTGAAGCATGACATTAAGTCTGG 983
Qy 514 AGGCTCTGGAAGAGTGTATCCCGCCCAACCATCAATGGCGCTTAAAGTCTAGGAAGCG 573
Db 984 AGGCTCTGGAAGAGTGTATCCCGCCCAACCATCAATGGCGCTTAAAGTCTAGGAAGCG 1043
Qy 574 GTGTGGTGGGCTTAGGGCAGGCGCAGACATACCCGAAAGTGGTGGATGATAC 633
Db 1044 GTGTGGTGGGCTTAGGGCAGGCGCAGACATACCCGAAAGTGGTGGATGATAC 1103
Qy 634 GCAAGGGCTGGATCGAAGCCCAAGACACTGGAAGCTGTGTGGTGGAGGGGCC 693
Db 1104 GCAAGGGCTGGATCGAAGCCCAAGACACTGGAAGCTGTGTGGTGGAGGGGCC 1163
Qy 694 GGCA 697
Db 1164 GGCA 1167
```

## RESULT 2

BM980506/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BM980506  
760 bp mRNA linear EST 21-FEB-2003  
UI-CF-EN1-add-a-08-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
UI-CF-EN1-add-a-08-0-UI 3', mRNA sequence.  
BM980506  
BM980506.1 GI:19602038  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 760)  
Ronald, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
889548  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul.mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

source

```
1..760
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-add-a-08-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"
```

## ORIGIN

Query Match

Best Local Similarity 99.4%; Pred. No. 0;

Matches 710; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Qy

240 GCTTCGCTGATGCACATTGGACCGGCTCTGGACTGGGTAGGGAGGAGGCGG 299

Db

734 GCTTCGCTGATGCACATTGGACCGGCTCTGGACTGGGTAGGGAGGAGGCGG 675

```
QY 300 GAATTGGGCGGAGGGCCAGGCGCTCGCGGACCCCGACTGCGGCTCCCGGTGGCCCGCGCA 359
Db 674 GAAATTGGGCGGAGGGCCAGGCGCTCGCGGACCCCGACTGCGGCTCCCGGTGGCCCGCGCA 615
QY 360 GCGGCTCCCGGTGGCGCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGG 419
Db 614 GCGGCTCCCGGTGGCGCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGG 555
QY 420 AGCTGTGCATGGGCGGCTGGCTAAAGCGCGTGGTTGGTTAGATGGCCAGCGGGAATTAA 479
Db 554 AGCTGTGCATGGGCGGCTGGCTAAAGCGCGTGGTTGGTTAGATGGCCAGCGGGAATTAA 495
QY 480 GTGTGCTCTGAAGAGCATGGACATAGTCTCGAGGGCTCTGAGAGTGTATCCCGGCC 539
Db 494 GTGTGCTCTGAAGAGCATGGACATAGTCTCGAGGGCTCTGAGAGTGTATCCCGGCC 435
QY 540 CCACCATCAATGGCGCTTAGGCTAGGAAGCGGGGTGTGGGTGGGCGCTTAGGGCGAGGC 599
Db 434 CCACCATCAATGGCGCTTAGGCTAGGAAGCGGGGTGTGGGTGGGCGCTTAGGGCGAGGC 375
QY 600 GCAGACATCCCGAAGTGTGGATTGTATACCGCAAGGGGTGGATCGAACCCCGCA 659
Db 374 GCAGACATCCCGAAGTGTGGATTGTATACCGCAAGGGGTGGATCGAACCCCGCA 315
QY 660 AGACACTGGAAGGCTGTGCTGCTGAGGAGGCGCGGCA-ATCCAGTGTGCTGCGGCTTT 718
Db 314 AGACACTGGAAGGCTGTGCTGCTGAGGAGGCGCGGCAAGTGTGCTGCGGCTTT 255
QY 719 ACAGGAAGAGTCCACCTTCT-TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGA 777
Db 254 ACAGGAAGAGTCCACCTTCT-TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGA 195
QY 778 TGGAGCTGCGGCGGCGAGAGTCTGCCAGTACAAAGCTGATTTGACCTGGGCGCT 837
Db 194 TGGAGCTGCGGCGGCGAGAGTCTGCCAGTACAAAGCTGATTTGACCTGGGCGCT 135
QY 838 CTGACATTCCTGATTCCTGCTGTCATCTCCAGCAAGTCTGTCCGCTGGCTGCTT 897
Db 134 CTGACATTCCTGATTCCTGCTGTCATCTCCAGCAAGTCTGTCCGCTGGCTGCTT 75
QY 898 CATCCACTCTCTCACTTCTGCTGCTCAGAGTAAATTCAGATCTGTGGTGC 951
Db 74 CATCCACTCTCTCACTTCTGCTGCTCAGAGTAAATTCAGATCTGTGGTGC 21

RESULT 3
BU620427/c 691 bp mRNA linear EST 23-SEP-2002
LOCUS
DEFINITION
UI-H-FL1-bfw-n-19-0-UI.s1 NCI CGAP FL1 Homo sapiens cDNA clone
BU620427
UI-H-FL1-bfw-n-19-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 691)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.
```

## FEATURES

```
Location/Qualifiers
1..691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FL1-bfw-n-19-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FL1"
(note="Organ: Chondrosarcoma; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site 1: Ecor I;
Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library
derived from a pool of mRNA obtained from 4 cell lines
from grade III chondrosarcoma tissues. The library was
constructed according to Bonaudo, Lemmon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
GAGTCCGCTG. The cell lines were provided by Dr. James
Martin from the University of Iowa.
TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
TAG L18-UI-H-FL1
TAG_SEQ=GAGTCCGCTG"
```

## ORIGIN

```
Query Match 34.6%; Score 343; DB 13; Length 691;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 304 TGGGCCCCAGGGCCAGGCGCTGCGGACCCCGACTGCGGCTCCCGGTGGCCCGCGCGC 363
Db 667 TGGGCCCCAGGGCCAGGCGCTGCGGACCCCGACTGCGGCTCCCGGTGGCCCGCGCGC 608
QY 364 CTCGCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGAGCT 423
Db 607 CTCGCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGAGCT 548
QY 424 GTGCATGGGCGCTGCGCTAAGGCCGTGGTTGGTTACGATTGGCCAGCGGAGCTTAAGTGT 483
Db 547 GTGCATGGGCGCTGCGCTAAGGCCGTGGTTGGTTACGATTGGCCAGCGGAGCTTAAGTGT 488
QY 484 TGCTCTCTGAAGAGCATGGACATTAGTCTGGAGGCTCTGGAAGAGTGTATCCCGCCCGC 543
Db 487 TGCTCTCTGAAGAGCATGGACATTAGTCTGGAGGCTCTGGAAGAGTGTATCCCGCCCGC 428
QY 544 CATCAATGGCGCTTAGGTCTAGGAAGCGGGGTGGGTGGGCGCTTAGGGCGAGCGCAG 603
Db 427 CATCAATGGCGCTTAGGTCTAGGAAGCGGGGTGGGTGGGCGCTTAGGGCGAGCGCAG 368
QY 604 ACATACCCCGAAGTGGTTGGAATTGATATACCGAAGGGGCTGATCGAACCCCGCAAGAC 663
Db 367 ACACACCCCGAAGTGGTTGGAATTGATATACCGAAGGGGCTGATCGAACCCCGCAAGAC 308
QY 664 ACTGGAAGGCTGTGTGGCTGAGGAGCGGCCCGCGCA 697
Db 307 ACTGGAAGGCTGTGTGGCTGAGGAGCGGCCCGCGCA 274
```

## RESULT 4

```
BU620427/c
LOCUS
DEFINITION
BX406138 Homo sapiens PETAL LIVER Homo sapiens cDNA clone
CS0DM010YD19 3-PRIME, mRNA sequence.
ACCESSION
BX406138
BX406138.1 GI:30648317
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
```



```

Db      522  ACGATTGGCCAGCGGACTTAAGTGTGTCTCTGAAGAGCATGACATTAGTCTGGAGG 463
Qy      518  TCTTGGAGAGTATCCCGCCGCCACCATCAATAGGCGTCTAGTCTAGAAAGCGGTGT 577
Db      462  TCTTGGAGAGTATCCCGCCGCCACCATCAATAGGCGTCTAGTCTAGAAAGCGGTGT 403
Qy      578  GGGTGGGCGCTTAGGGGAGGCGGAGACATACACCCGAAAGTGGTGGATTGTATACCGCAA 637
Db      402  GGGTGGGCGCTTAGGGGAGGCGGAGACACACCCGAAAGTGGTGGATTGTATACCGCAA 343
Qy      638  GGGGCTGGATCGAACCCGCCCAAGACACTGGAGAGGTGTGTGCTGAGAGGCGCCCGGCA 697
Db      342  GGGGCTGGATCGAACCCGCCCAAGACACTGGAGAGGTGTGTGCTGAGAGGCGCCCGGCA 283
Qy      698  -ATCCAGTGTGTCTGGGCGCTTACAGGAAAGAGCTCCACCTTCT-TGGAGTGTGCAGATG 755
Db      282  GATCCAGTGTGTCTGGGCGCTTACAGGAAAGAGCTCCACCTTCTTGGAGTGTGCAGATG 223
Qy      756  CGATCTAGGTGTGTCCACCGATGGAGCTGGGCGGCGGCGAGATGCTGCCCGAGTACAA 815
Db      222  CGATCTAGGTGTGTCCACCGATGGAGCTGGGCGGCGGCGAGATGCTGCCCGAGTACAA 163
Qy      816  AGCTGATTTCCGACTGGGCGCTCGGACTTCCCTGATTTCTCTGCTTGCATCTCCACCAA 875
Db      162  AGCTGATTTCCGACTGGGCGCTCGGACTTCCCTGATTTCTCTGCTTGCATCTCCACCAA 103
Qy      876  GTCTGTCTCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTGCATGTAATAATT 935
Db      102  GTCTGTCTCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTGCATGTAATAATT 43
Qy      936  GCAAGATCTGTGTGC 951
Db      42  GCAAGATCTGTGTGC 27

```

## RESULT 6

```

BU684594/c
LOCUS      694 bp mRNA linear EST 07-OCT-2002
DEFINITION UI-CF-EN1-act-a-22-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
            UI-CF-EN1-act-a-22-0-UI 3', mRNA sequence.
ACCESSION  BU684594
VERSION     BU684594.1 GI:23537704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 694)
AUTHORS     Bonaldo M.F., Lennon G. and Soares M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: McCray, PB
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            POLYA-Yes.
            Location/Qualifiers
            1..694
            /organism="Homo sapiens"

```

## FEATURES

source

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-act-a-22-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site_1: Sc0r 1; Site_2: Not 1;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into p773-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

```

## ORIGIN

```

Query Match      32.6%; Score 323; DB 13; Length 694;
Best Local Similarity 99.5%; Fred. No. 0;
Matches 613; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
Qy      338  TCGCGCTCCCGGTGGCGCCCGCAGCGCCCTCCCGGTGGCGCCCTGGAGTGCAGGCTTACCGTC 397
Db      647  TCGCGCTCCCGGTGGCGCCCGCAGCGCCCTCCCGGTGGCGCCCTGGAGTGCAGGCTTACCGTC 588
Qy      398  CGAGATCTCGCAACTGGCGAGCTGTGCATGGCGCGTGGCTAAGCGCGTGGTTGGTT 457
Db      587  CGAGATCTCGCAACTGGCGAGCTGTGCATGGCGCGTGGCTAAGCGCGTGGTTGGTT 528
Qy      458  ACGATTGGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGG 517
Db      527  ACGATTGGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGG 468
Qy      518  TCTCGAAGAGTGTATCCCGCCCGCCACCATCAATGGCGCTTAGGTCTAGGAAGCGGGTGT 577
Db      467  TCTCGAAGAGTGTATCCCGCCCGCCACCATCAATGGCGCTTAGGTCTAGGAAGCGGGTGT 408
Qy      578  GGGTGGGCGCTTAGGGCGAGCGGAGACATACCCCGAGTGGTTGGATTGTATACCGCAA 637
Db      407  GGGTGGGCGCTTAGGGCGAGCGGAGACACACCCCGAAGTGGTTGGATTGTATACCGCAA 348
Qy      638  GGGGCTGGATCGAACCCCGCCCAAGACACTGGAAAGCTGTGTGGCTGAGGAGGGCCCGGCA 697
Db      347  GGGGCTGGATCGAACCCCGCCCAAGACACTGGAAAGCTGTGTGGCTGAGGAGGGCCCGGCA 288
Qy      698  -ATCCAGTGTGTGTGGCGCTTACAGGAAAGAGCTCCACCTTCT-TGAGTGTGCAGATG 755
Db      287  GATCCAGTGTGTGTGGCGCTTACAGGAAAGAGCTCCACCTTCTCTGAGTGTGCAGATG 228
Qy      756  CGATCTAGGTGTGTCCACCGATGGAGCTCGGCGCGGCGAGATGCTGCCCGAGTACAA 815
Db      227  CGATCTAGGTGTGTCCACCGATGGAGCTCGGCGCGGCGAGATGCTGCCCGAGTACAA 168
Qy      816  AGCTGATTTGAGACTGGGCGCTCTGGACTTCCCTGATTTCTCTGCTTGCATCTCCAGCAA 875
Db      167  AGCTGATTTGAGACTGGGCGCTCTGGACTTCCCTGATTTCTCTGCTTGCATCTCCAGCAA 108
Qy      876  GTCTGTCTCCGCTTGGCTTGCATCCACTCTCTCACTTCTCTGCTTGCATGTAATAATT 935
Db      107  GTCTGTCTCCGTTGGCTTGCATCCACTCTCTCACTTCTCTGCTTGCATGTAATAATT 48
Qy      936  GCAAGATCTGTGTGC 951

```

```

Db      47 GCAGATCTGTGGTGC 32
|||||
RESULT 7
CD742581/c
LOCUS      623 bp mRNA linear EST 26-JUN-2003
DEFINITION UI-H-PT2-bj1-i-10-0-UI.s1 NCI CGAP_FT2 Homo sapiens cDNA clone
            UI-H-PT2-bj1-i-10-0-UI 3', mRNA sequence.
ACCESSION  CD742581
VERSION     CD742581.1 GI:32293431
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 623)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Dr. Gary W. Hunninghake, U of I
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/cgap.html
            Seq primer: M13 FORWARD
            POLYA=yes.
            Location/Qualifiers
                1..623
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-H-PT2-bj1-i-10-0-UI"
                /tissue_type="Aveolar Macrophage"
                /dev_stage="Adult"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="NCI CGAP FT2"
                /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                modified polylinker; Site 1: EcoR I; Site 2: Not I;
                NCI-CGAP_FT2 is a subtracted cDNA library constructed from
                a pool of 81 RNA samples from Alveolar Macrophages
                challenged with different treatments. The library was
                substracted according to Bonaldo, Lennon and Soares, Genome
                Research, 6:791-806, 1996. The tissue was provided by Dr.
                Gary W. Hunninghake of the University of Iowa.
                TAG_TISSUE=Human Lung Aveolar Macrophage
                TAG_LIB=UI-H-PT2
                TAG_SEQ=GGCCATGCGC"

ORIGIN
Query Match      31.5%; Score 312; DB 14; Length 623;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 602; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY      343 CTCGGGTGCGCCGAGCGCCCTCCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGA 402
Db      623 CTCGGGTGCGCCGAGCGCCCTCCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGA 564
QY      403 TCGTCGCACTGGCGAGCTGTCATGGGCGTGGCTAAGCGCGGTGGTTGGTTACGAT 462
Db      563 TCGTCGCACTGGCGAGCTGTCATGGGCGTGGCTAAGCGCGGTGGTTGGTTACGAT 504
QY      463 TGCCACGCGGACTTAAGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGGTCTCTG 522
Db      503 TGCCACGCGGACTTAAGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGGTCTCTG 444
QY      523 GAGAGTATCCCGCCACCATCAATGCGCTTAGCTCTAGGAGCGGGTGGGTG 582
Db      443 GAAGAGTGATCCCGCCGCCACCATCAATGCGCTTAGCTCTAGGAGCGGGTGGGTG 384

```

```

QY      583 GGGCTTAGGGCGGCGCAGACATACCCGAACTGGTTGGATTGTATACCGCAAGGGC 642
Db      383 GGGCTTAGGGCGGCGCAGACACACCCGAACTGGTTGGATTGTATACCGCAAGGGC 324
QY      643 TGGATCGAAACCCCAAGACACTGGAAGGCTGTGTGCTGAGGAGGCGCCGGCA-ATCC 701
Db      323 TGGATCGAAACCCCAAGACACTGGAAGGCTGTGTGCTGAGGAGGCGCCGGCAGATCC 264
QY      702 AGTGTGCTGTGGGCTTTACAGGAAAGAGCTCCACTTCT- TGGAGTGTGAGATGCCATC 760
Db      263 AGTGTGCTGTGGGCTTTACAGGAAAGAGCTCCACTTCTCTGGAGTGTGAGATGCCATC 204
QY      761 TAGGTGTCTCCACCGCATGGGAGCTGGGGCGGGGCGAGATGCTGCCCCAGTACAAAGCTG 820
Db      203 TAGGTGTCTCCACCGCATGGGAGCTGGGGCGGGGCGAGATGCTGCCCCAGTACAAAGCTG 144
QY      821 ATTTGGACCTGGGCGCTCTGGACTTCCCTGATTCCTGTGCTTGCATCTCCAGCAAAAGTCT 880
Db      143 ATTTGGACCTGGGCGCTCTGGACTTCCCTGATTCCTGTGCTTGCATCTCCAGCAAAAGTCT 84
QY      881 CTCGGTGGCTGCTTCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTGCAAG 940
Db      83 CTCGGTGGCTGCTTCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTGCAAG 24
QY      941 ATCTG 945
Db      23 ATCTG 19

RESULT 8
BU617331/c
LOCUS      632 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-DF0-bep-n-09-0-UI.s1 NCI CGAP_DF0 Homo sapiens cDNA clone
            UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.
ACCESSION  BU617331
VERSION     BU617331.1 GI:23283539
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 632)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=yes.
            Location/Qualifiers
                1..632
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-H-DF0-bep-n-09-0-UI"
                /tissue_type="Subchondral Bone"
                /dev_stage="Adult"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="NCI CGAP_DF0"
                /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                modified polylinker; Site 1: EcoR I; Site 2: Not I;
                NCI CGAP_DF0 is a cDNA library containing the following
                tissue(s): Subchondral Bone. The library was constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed

```

FEATURES  
source



with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.

TAG\_TISSUE=chondral bone  
TAG\_LIB=UI-H-DF0  
TAG\_SEQ=GTTAAGCGTC

## ORIGIN

Query Match	31.1%;	Score 308;	DB 13;	Length 632;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 358;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

QY 339 GCGCTCCCGTGGCGCCCGAGCGCTCCCGGTGGCCCTGAGTGCAGTCTTACCGTCC 398

Db 632 GCGCTCCCGTGGCGCCCGAGCGCTCCCGGTGGCCCTGAGTGCAGTCTTACCGTCC 573

QY 399 GAGATCGTCCGCAACTGGCGGAGCTGTGTCATGGGCGTGGCTAAGGCGGTGTTTGTGA 458

Db 572 GAGATCGTCCGCAACTGGCGGAGCTGTGTCATGGGCGTGGCTAAGGCGGTGTTTGTGA 513

QY 459 CGATTGGCCAGCGGACTTAAGTGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGGT 518

Db 512 CGATTGGCCAGCGGACTTAAGTGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGGT 453

QY 519 CCTGGAAGAGTGATCCCGCCCGCCACCATCAATCGCGCTTAGGCTAGGAGCGGGTGTG 578

Db 452 CCTGGAAGAGTGATCCCGCCCGCCACCATCAATCGCGCTTAGGCTAGGAGCGGGTGTG 393

QY 579 GGTGGGCGCTTAGGCGGAGCGCGAGACATACCCGAAAGTGGTGGATTGTATACCGCAAG 638

Db 392 GGTGGGCGCTTAGGCGGAGCGCGAGACATACCCGAAAGTGGTGGATTGTATACCGCAAG 333

QY 639 GGGCTGGATCGAATCCCGCCCAAGACACTGGAAGCTGTGGCTGAGAGGGCGCCGCGCA 697

Db 332 GGGCTGGATCGAATCCCGCCCAAGACACTGGAAGCTGTGGCTGAGAGGGCGCCGCGCA 274

RESULT 9  
B0447041/c  
LOCUS  
DEFINITION UI-H-EUI-bac-p-06-0-UI.s1 NCI CGAP Ctl Homo sapiens linear EST 29-MAY-2002  
B0447041  
UI-H-EUI-bac-p-06-0-UI 3', mRNA sequence.  
B0447041  
EST.  
B0447041.1 GI:21250153  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 628)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

Location/Qualifiers  
1..628  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

## source

Contact: McCray, PB

/clone="UI-H-EUI-bac-p-06-0-UI"  
/tissue\_type="Osteoarthritic Cartilage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP Ctl1"  
/note="Organ: Knee; Vector: p7773-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site 2: Not I;  
NCI CGAP Ctl1 is a normalized cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTATCACGCT.  
TAG\_TISSUE=osteoarthritic cartilage  
TAG\_LIB=UI-H-EUI  
TAG\_SEQ=GTATCACGCT"

## ORIGIN

Query Match	27.9%;	Score 276;	DB 13;	Length 628;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 326;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

QY 371 TGGCCCTGGAGTGCAGTCTTACCGTCCGAGATCGTCGCAACTGGCGAGTGTGCATG 430

Db 601 TGGCCCTGGAGTGCAGTCTTACCGTCCGAGATCGTCGCAACTGGCGAGTGTGCATG 542

QY 431 GGGCGTGGCTAAGCGCGTGGTTGGTTAGATTGGCCAGCGGAGCTTAAGTGTGTCTCT 490

Db 541 GGGCGTGGCTAAGCGCGTGGTTGGTTAGATTGGCCAGCGGAGCTTAAGTGTGTCTCT 482

QY 491 GAAGAGCATGGACATTAGTCTGGAGGGTCTCTGGAAGAGTGATCCCGCCGCCACCATCAA 550

Db 481 GAAGAGCATGGACATTAGTCTGGAGGGTCTCTGGAAGAGTGATCCCGCCGCCACCATCAA 422

QY 551 TGGCCCTTAGGCTTAGGAAGCGGTGTGGTGGGCGCTTAGGGCGAGGCGGAGATACC 610

Db 421 TGGCGCTTAGGCTTAGGAAGCGGTGTGGTGGGCGCTTAGGGCGAGGCGGAGACACACC 362

QY 611 CGGAAGTGGTGGATTGTATATACCGCAAGGGGTGCGATCGAACCCGCCAAAGACACTGGAA 670

Db 361 CGGAAGTGGTGGATTGTATATACCGCAAGGGGTGCGATCGAACCCGCCAAAGACACTGGAA 302

QY 671 GGCTGTGGCTGAGGAGGGCGCCGCA 697

Db 301 GGCTGTGGCTGAGGAGGGCGCCGCA 275

## RESULT 10

CA313142/c

LOCUS

DEFINITION UI-CF-FNO-aex-n-23-0-UI.s1 UI-CF-FNO Homo sapiens linear EST 04-NOV-2002

CA313142

UI-CF-FNO-aex-n-23-0-UI 3', mRNA sequence.

CA313142.1 GI:24531240

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 633)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

COMMENT

Tue Jul 13 16:16:20 2004

McCrack Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.

Location/Qualifiers  
1. .633  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-FNO-aex-n-23-0-UI"  
/tissue\_type="Human Lung Epithelial cells"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-FNO"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-FNO is a subcloned cDNA library derived from two  
normalized Human lung epithelial cell libraries (EN1 and  
DUI). The library was subcloned according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. For additional information, contact:  
bento-soares@uiowa.edu  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG LIB=UI-CF-FNO  
TAG\_SEQ=CTGCTCAGGT"

## FEATURES

source

## ORIGIN

Query Match 27.2%; Score 269; DB 14; Length 633;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 609; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 341 GCCTCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCCAGTCTTACCGTCCGA 400  
DB 633 GCCTCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCCAGTCTTACCGTCCGA 574

QY 401 GATCGTCCGCACTGGGCGAGCTGTGATGGGCGTGGCTAAGGCGCTGTTGGTTACG 460  
DB 573 GATCGTCCGCACTGGGCGAGCTGTGATGGGCGTGGCTAAGGCGCTGTTGGTTACG 514

QY 461 ATTGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGACATTAAGTCTGGAGGTCC 520  
DB 513 ATTGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGACATTAAGTCTGGAGGTCC 454

QY 521 TGGAGAGTGTATCCCGCCACCATCAATGGCGCTTAGTCTAGGAAGCGGTGTGGG 580  
DB 453 TGGAGAGTGTATCCCGCCACCATCAATGGCGCTTAGTCTAGGAAGCGGTGTGGG 394

QY 581 TGGGSCCTTAGGGCGAGCGCAGACATACCCCGAAGTGTGTGATTGTATACCGCAAGG 640  
DB 393 TGGGSCCTTAGGGCGAGCGCAGACATACCCCGAAGTGTGTGATTGTATACCGCAAGG 334

QY 641 GCTGGATCGAACCCCGCAAGACACTGAAGGCTGTGGCTGAGGAGGCGCGGCA-AT 699  
DB 333 GCTGGATCGAACCCCGCAAGACACTGAAGGCTGTGGCTGAGGAGGCGCGGCAAT 274

QY 700 CCAGTGTGTCTGGGCTTTACAGGAAGAGCTCCACCTTC-TTGGAGTGTGCAGATCGGA 758  
DB 273 CCAGTGTGTCTGGGCTTTACAGGAAGAGCTCCACCTTC-TTGGAGTGTGCAGATCGGA 214

QY 759 TCTAGTGTGTCCACCGATGGAGCTGCGGGCGGCGAGATGCTGCCCGCAGTACAAGC 818  
DB 213 TCTAGTGTGTCCACCGATGGAGCTGCGGGCGGCGAGATGCTGCCCGCAGTACAAGC 154

QY 819 TGATTGGACCTGGGGCTCTGGGACTTCCCTGATTCTCTGCTTGCATCTCCAGCAAGTC 878  
DB 153 TGATTGGACCTGGGGCTCTGGGACTTCCCTGATTCTCTGCTTGCATCTCCAGCAAGTC 94

QY 879 CTGTCCCGTGGCTGGCTTATCCACTCTCTCACTCTCTGCTTGCCTTCAGAGTAAATTCGA 938  
DB 93 CTGTCCCGTGGCTGGCTTATCCACTCTCTCACTCTCTGCTTGCCTTCAGAGTAAATTCGA 34

QY 939 AGATCTGTGGTGC 951  
DB 33 AGATCTGTGGTGC 21

RESULT 11  
AA687318/c  
LOCUS AA687318  
DEFINITION nvl81e01.sl NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1234296 3',  
mRNA sequence.  
ACCESSION AA687318  
VERSION AA687318.1 GI:2675509  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 955)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb@mail.nih.gov  
unknown library type  
Insert Length: 966 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 361.

FEATURES  
source  
1. .395  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1234296"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/note="Vector: pT73-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I-Oligo (dT) primer  
15'-TGTTACCAATCTGAAGTGGAGCGCCCTCATTTTCTTTTCTTTT-3',  
J. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 22.7%; Score 225; DB 9; Length 395;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 TGGAGTGTGCAGATCGATCTAGGTGTGTCCACCGGAGGTGGCGCGGCGAGAT 800  
DB 225 TGGAGTGTGCAGATCGATCTAGGTGTGTCCACCGGAGGTGGCGCGGCGAGAT 166

QY 801 GCTGCCCGCAGTACAAGCTGATTTCGAGCTTGGGGCTCTGGACTTCCTGATTCCTGCT 860  
DB 165 GCTGCCCGCAGTACAAGCTGATTTCGAGCTTGGGGCTCTGGACTTCCTGATTCCTGCT 106

```

QY 861 TGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTGC 920
Db 105 TGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTGC 46
QY 921 CTTTCAGAGTAAATTCGAAGATCTGTGGTGCACAAAAA 965
Db 45 CTTTCAGAGTAAATTCGAAGATCTGTGGTGCACAAAAA 1

RESULT 12
BM672504/c
LOCUS
DEFINITION
  UI-E-CQ1-ada-e-09-0-UI.s1 UI-E-CQ1 Homo sapiens cDNA clone
BM672504
ACCESSION
  BM672504
VERSION
  BM672504.1 GI:118982402
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 300)
  Bonaudo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 5 (9), 791-806 (1996)
  97044477
  8889548
  Contact: Soares, MB
  Coordinated Laboratory for Computational Genomics
  University of Iowa
  375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
  Tel: 319 335 8250
  Fax: 319 335 9565
  Email: bento-soares@uiowa.edu
  Tissue Procurement: Dr. Gregg Hageman
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  The following repetitive elements were found in this cDNA
  sequence: 1-21, >AT rich#Low_complexity (matched complement)
  Seq primer: M13 Forward
  POLYA=Yes.

FEATURES
  source
  1..300
    Location/Qualifiers
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-E-CQ1-ada-e-09-0-UI"
    /tissue_type="optic nerve"
    /dev_stage="adult"
    /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-E-CQ1"
    /notes="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-E-CQ1 is a normalized cDNA library containing the
    following tissue(s): optic nerve. The library was
    constructed according to Bonaudo, Lennon and Soares,
    Genome Research, 5:791-806, 1996. First strand cDNA
    synthesis was primed with an oligo-dT primer containing a
    Not I site. Double stranded cDNA was ligated to an EcoR I
    adaptor, digested with Not I, and cloned directionally
    into pTV73-Pac vector. The oligonucleotide used to prime
    the synthesis of first-strand cDNA contains a library tag
    sequence that is located between the Not I site and the
    (dfr)18 tail. The sequence tag for this library is
    CCAATTAAGTG. This library was created for the program, Gene
    Discovery in the Visual System, supported by National Eye
    Institute (NEI).
    TAG_TISSUE=human
    TAG_LIB=UI-E-CQ1

```

## TAG\_SEQ=CCATTAAGTG

## ORIGIN

```

Query Match      21.4%; Score 212; DB 12; Length 300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 740 TTGGAGTGTGCAGATCGCATCTAGGTGTCTCCACCCGATGGAGCTGGGGCCGGCAGA 799
Db 233 TTGGAGTGTGCAGATCGCATCTAGGTGTCTCCACCCGATGGAGCTGGGGCCGGCAGA 174
QY 800 TGCCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCCCTCTGGACTTCCTGATCTCTGC 859
Db 173 TGCCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCCCTCTGGACTTCCTGATCTCTGC 114
QY 860 TTGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTG 919
Db 113 TTGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTG 54
QY 920 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 951
Db 53 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 22

```

## RESULT 13

## BM687366

## LOCUS

## DEFINITION

UI-E-CQ1-ada-e-09-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone

## ACCESSION

## BM687366

## VERSION

## BM687366.1

## KEYWORDS

## EST.

## SOURCE

## Homo sapiens (human)

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## 1 (bases 1 to 462)

## Bonaudo,M.F., Lennon,G. and Soares,M.B.

## Normalization and subtraction: two approaches to facilitate gene

## discovery

## Genome Res. 6 (9), 791-806 (1996)

## 97044477

## 8889548

## Contact: Soares, MB

## Coordinated Laboratory for Computational Genomics

## University of Iowa

## 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

## Tel: 319 335 8250

## Fax: 319 335 9565

## Email: bento-soares@uiowa.edu

## Tissue Procurement: Dr. Gregg Hageman

## cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

## cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

## DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

## Clone Distribution: Researchers may obtain clones from Research

## Genetics (www.resgen.com).

## Seq primer: M13 Reverse

## Location/Qualifiers

## 1..462

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="UI-E-CQ1-ada-e-09-0-UI"

## /tissue\_type="optic nerve"

## /dev\_stage="adult"

## /lab\_host="DHI0B (Life Technologies) (T1 phage resistant)"

## /clone\_lib="UI-E-CQ1"

## /notes="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a

## modified polylinker; Site 1: EcoR I; Site 2: Not I;

## UI-E-CQ1 is a normalized cDNA library containing the

## following tissue(s): optic nerve. The library was

## constructed according to Bonaudo, Lennon and Soares,

## Genome Research, 5:791-806, 1996. First strand cDNA

## synthesis was primed with an oligo-dT primer containing a

## Not I site. Double stranded cDNA was ligated to an EcoR I

## adaptor, digested with Not I, and cloned directionally

## into pTV73-Pac vector. The oligonucleotide used to prime

## the synthesis of first-strand cDNA contains a library tag

## sequence that is located between the Not I site and the

## (dfr)18 tail. The sequence tag for this library is

## CCAATTAAGTG. This library was created for the program, Gene

## Discovery in the Visual System, supported by National Eye

## Institute (NEI).

## TAG\_TISSUE=human

## TAG\_LIB=UI-E-CQ1

synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7733-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCAATAGTG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Query Match 21.3%; Score 211; DB 12; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 TGGAGTGTGCAGATGGCATAGTGTGTCACCCGATGGAGCTGCGGGCCGGGCAGAT 800  
 |||||  
 DB 252 TGGAGTGTGCAGATGGCATAGTGTGTCACCCGATGGAGCTGCGGGCCGGGCAGAT 311  
 |||||

QY 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCTCTGGACTTCCCTGATCTCTGCT 860  
 |||||

DB 312 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCTCTGGACTTCCCTGATCTCTGCT 371  
 |||||

QY 861 TGCATCTCCAGCAAGTCTGCTCCCTGGCTGGCTTCATCCACTCTCTCACTTCTCTGCG 920  
 |||||

DB 372 TGCATCTCCAGCAAGTCTGCTCCCTGGCTGGCTTCATCCACTCTCTCACTTCTCTGCG 431  
 |||||

QY 921 CTTCAGAGTAAATTCGAAGATCTGTGGTGC 951  
 |||||

DB 432 CTTCAGAGTAAATTCGAAGATCTGTGGTGC 462  
 |||||

## RESULT 14

BF509528/c  
 LOCUS 480 bp mRNA linear EST 06-DEC-2000  
 DEFINITION UI-H-B14-aos-f-10-0-UI.s1 NCI\_CGAP\_Sub8 Homo sapiens cDNA clone IMAGE:3086731 3', mRNA sequence.

ACCESSION BF509528  
 VERSION BF509528.1 GI:11592826  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 480)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library Preparation: M.B. Soares Lab clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

source  
 1. 480  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3086731"  
 /lab\_hosts="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP Sub8"  
 /notes="Vector: p7733D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; NCI\_CGAP Sub8 is a subtracted library derived from NCI\_CGAP Sub5. The NCI\_CGAP Sub8 library had 2.5 million recombinants. A

single-stranded DNA preparation of NCI CGAP Subs was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI\_CGAP\_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE ids 272869-2731190; 25% of the driver population), and NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described (Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG\_TISSUE=lung  
 TAG\_LIB=NCI\_CGAP\_Lul9  
 TAG\_SEQ=GACAGC"

Query Match 21.3%; Score 211; DB 10; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 TGGAGTGTGCAGATGGCATAGTGTGTCACCCGATGGAGCTGCGGGCCGGGCAGAT 800  
 DB 230 TGGAGTGTGCAGATGGCATAGTGTGTCACCCGATGGAGCTGCGGGCCGGGCAGAT 171  
 QY 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCTCTGGACTTCCCTGATCTCTGCT 860  
 DB 170 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCTCTGGACTTCCCTGATCTCTGCT 111  
 QY 861 TGCATCTCCAGCAAGTCTGCTCCCTGGCTGGCTTCATCCACTCTCTCACTTCTCTGCG 920  
 DB 110 TGCATCTCCAGCAAGTCTGCTCCCTGGCTGGCTTCATCCACTCTCTCACTTCTCTGCG 51  
 QY 921 CTTCAGAGTAAATTCGAAGATCTGTGGTGC 951  
 DB 50 CTTCAGAGTAAATTCGAAGATCTGTGGTGC 20

## RESULT 15

BF002915/c  
 LOCUS 367 bp mRNA linear EST 06-OCT-2000  
 DEFINITION 795Ch12.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:3309959 3', mRNA sequence.

ACCESSION BF002915  
 VERSION BF002915.1 GI:10703190  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 367)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
 Seq primer: -40UP from Gibco.

## FEATURES

source  
 1. 367  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

```
/clone="IMAGE:3309959"  
/sex="male"  
/dev_stage="adult"  
/lab_host="DH10B"  
/clone_lib="NCI_CGAP_Pr28"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI_CGAP_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "
```

## ORIGIN

```
Query Match      21.2%; Score 210; DB 10; Length 367;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 741 TGGAGTGTGCAGATGCGATCTAGGTGTGTCACCCGATGGGAGCTGGGGCCGGGCGAGAT 800  
    |||||  
Db 214 TGGAGTGTGCAGATGCGATCTAGGTGTGTCACCCGATGGGAGCTGGGGCCGGGCGAGAT 155  
    |||||  
Qy 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCCTCTGGACTTCCTGATTCTCTGCT 860  
    |||||  
Db 154 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCCTCTGGACTTCCTGATTCTCTGCT 95  
    |||||  
Qy 861 TGCATCTCCAGCAAGTCCCTGTCCTGCTGCTTCATCCACTCTCTCATTCTCTGCT 920  
    |||||  
Db 94 TGCATCTCCAGCAAGTCCCTGTCCTGCTGCTTCATCCACTCTCTCATTCTCTGCT 35  
    |||||  
Qy 921 CTTCAGAGTAAATTCGAAGATCTGGGTG 950  
    |||||  
Db 34 CTTCAGAGTAAATTCGAAGATCTGGGTG 5  
    |||||
```

Search completed: July 10, 2004, 15:29:05  
Job time : 3115 secs

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model  
Run on: July 10, 2004, 11:58:39 ; Search time 4183 Seconds  
(without alignments)  
10458.093 Million cell updates/sec

Title: US-09-972-032-1  
Perfect score: 990  
Sequence: 1 ggaatgttcttcgaggccaa.....aaaatgagcgccgaagtt 990

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pin.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	579	58.5	182230	9	AC135050	AC135050 Homo sapi
2	508	51.3	185476	2	AC135044	AC135044 Homo sapi
C 3	79	8.0	96975	9	AC004975	AC004975 Homo sapi
C 4	66	6.7	129758	9	AC118269	AC118269 Homo sapi
C 5	66	6.7	138536	9	AC116914	AC116914 Homo sapi
6	66	6.7	140210	2	AC002993	AC002993 Homo sapi
7	66	6.7	209885	2	AC011189	AC011189 Homo sapi
8	63	6.4	143065	9	HS20208	AL031848 Human DNA
C 9	60	6.1	113872	9	HS57629	Z95115 Human DNA
C 10	59	6.0	44174	9	AC004645	AC004645 Homo sapi
C 11	59	6.0	45454	9	AF157412	AL157412 Human DNA
C 12	59	6.0	122029	9	AF212832	AF212832 Homo sapi
C 13	59	6.0	123291	9	AC104070	AC104070 Homo sapi
C 14	59	6.0	148963	2	BX571818	BX571818 Homo sapi
15	59	6.0	159619	2	AC012440	AC012440 Homo sapi
16	59	6.0	166843	2	AC090821	AC090821 Homo sapi
17	59	6.0	169003	2	AC026071	AC026071 Homo sapi
18	59	6.0	169893	2	AC021165	AC021165 Homo sapi
C 19	59	6.0	181558	2	AC068301	AC068301 Homo sapi
20	59	6.0	189207	9	AC103706	AC103706 Homo sapi
21	59	6.0	190814	9	AC006111	AC006111 Homo sapi
22	59	6.0	199776	9	AC011499	AC011499 Homo sapi
C 23	59	6.0	213056	9	AC103590	AC103590 Homo sapi
C 24	58	5.9	105574	2	AC013308	AC013308 Homo sapi
25	58	5.9	111388	9	AC016655	AC016655 Homo sapi
26	58	5.9	115487	2	AC027344	AC027344 Homo sapi
27	58	5.9	118447	2	AC093508	AC093508 Homo sapi
28	58	5.9	124752	9	AC010419	AC010419 Homo sapi
29	58	5.9	132117	9	AC092463	AC092463 Homo sapi
C 30	58	5.9	160956	2	AC013637	AC013637 Homo sapi
31	58	5.9	161144	9	AC074142	AC074142 Homo sapi
32	58	5.9	170000	2	AC004524	AC004524 Homo sapi
33	58	5.9	183003	9	AC106045	AC106045 Homo sapi
34	58	5.9	197087	9	AC009964	AC009964 Homo sapi
35	58	5.9	200000	2	AC004604	AC004604 Homo sapi
36	58	5.9	210000	2	AC004555	AC004555 Homo sapi
37	57	5.8	115663	9	AC093029	AC093029 Homo sapi
38	57	5.8	138063	9	AC092473	AC092473 Homo sapi
C 39	57	5.8	156831	2	AC074268	AC074268 Homo sapi
C 40	57	5.8	162104	9	AC026325	AC026325 Homo sapi
C 41	57	5.8	213546	2	AC136602	AC136602 Homo sapi
C 42	56	5.7	164028	9	AC004890	AC004890 Homo sapi
C 43	56	5.7	167679	9	AC007570	AC007570 Homo sapi
44	56	5.7	186829	9	BX640519	BX640519 Human DNA
45	56	5.7	270178	2	BX572623	BX572623 Homo sapi

ALIGNMENTS

RESULT 1  
AC135050  
LOCUS AC135050 182230 bp DNA linear PRI 27-FEB-2003  
DEFINITION Homo sapiens chromosome 16 clone RP11-196G11, complete sequence.  
ACCESSION AC135050  
VERSION AC135050.3 GI:28570306  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 182230)  
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
TITLE Direct Submission

```

JOURNAL
REFERENCE
  2 (bases 1 to 182230)
  DOE Joint Genome Institute.
AUTHORS
  Direct Submission
JOURNAL
  Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
  Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
  3 (bases 1 to 182230)
  DOE Joint Genome Institute.
AUTHORS
  Direct Submission
JOURNAL
  Submitted (22-OCT-2002) Production Sequencing Facility, DOE Joint
  Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
  4 (bases 1 to 182230)
  DOE Joint Genome Institute, Stanford Human Genome Center and Los
  Alamos National Laboratory.
AUTHORS
  Direct Submission
JOURNAL
  Submitted (27-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
COMMENT
  On Feb 27, 2003 this sequence version replaced gi:24211094.
  Draft Sequence Produced by DOE Joint Genome Institute
  www.jgi.doe.gov
  Finishing Completed at Stanford Human Genome Center and Los Alamos
  National Laboratory
  www.shgc.stanford.edu
  Quality: Phrap Quality >=40 99.7% of Sequence;
  Estimated Total Number of Errors is 0.3.
FEATURES
  source
    1. .182230
      /location=Qualifiers
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosomes="16"
      /clone="RP11-196G11"
ORIGIN
  Query Match          58.5%; Score 579; DB 9; Length 182230;
  Best Local Similarity 99.8%; Pred. No. 6.5e-312;
  Matches 819; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
  QY 133 GGCCCTGTGTGCTGCGCGCTGCCCGCGCTCCAGCTGCCAGCTGCGCGGCTCAGTC 192
  DB 98019 GGCCCTGTGTGCTGCGCGCTGCCCGCGCTCCAGCTGCCAGCTGCGCGGCTCAGTC 98078
  QY 193 CCCTGTGCATGTGTGGAGACCGCGTCGCTAGCGTGCATGTGCTTCCGTGATGC 252
  DB 98079 CCCTGTGCATGTGTGGAGACCGCGTCGCTAGCGTGCATGTGCTTCCGTGATGC 98138
  QY 253 ACATTGGACCGCGCTCTGACCTGCGCTAGCGGAAGGCGAGGCGGGAATTGGGCCCGA 312
  DB 98139 ACATTGGACCGCGCTCTGACCTGCGCTAGCGGAAGGCGAGGCGGGAATTGGGCCCGA 98198
  QY 313 GGCCAGGCTGCTCCGACCCCGGCTGCGCTCCCGGTCGCCCGGCGCTCCCGGTC 372
  DB 98199 GGCCAGGCTGCTCCGACCCCGGCTGCGCTCCCGGTCGCCCGGCGCTCCCGGTC 98258
  QY 373 GCCTCGAGTCAGGCTTACCGTCGAGATCGTCGCACTGGCGGCTGTCATGGG 432
  DB 98259 GCCTCGAGTCAGGCTTACCGTCGAGATCGTCGCACTGGCGGCTGTCATGGG 98318
  QY 433 CGTGCTAAGCGCGTGTGTTGTTACGATTCGCGGCGGCTTAAAGTGTGTTCTCGA 492
  DB 98319 CGTGCTAAGCGCGTGTGTTGTTACGATTCGCGGCGGCTTAAAGTGTGTTCTCGA 98378
  QY 493 AGAGCATGACATTAGTCGAGGCTTCGGAAGTATCCCGCCCGCCACCATCAATG 552
  DB 98379 AGAGCATGACATTAGTCGAGGCTTCGGAAGTATCCCGCCCGCCACCATCAATG 98438
  QY 553 GCGCTTAGGTCTAGGAAGCGGCTGTGGTGGGCGCTTAGGGCGAGCGCAGACATACCC 612
  DB 98439 GCGCTTAGGTCTAGGAAGCGGCTGTGGTGGGCGCTTAGGGCGAGCGCAGACATACCC 98498
  QY 613 GAAGTGGTGGATTGTATACCGCAAGCGGCTGATCGAACCCCGCCCAAGACACTGGAAG 672
  DB 98499 GAAGTGGTGGATTGTATACCGCAAGCGGCTGATCGAACCCCGCCCAAGACACTGGAAG 98558

```

```

QY 673 CTGTGTGGCTGAGAGGGCCCGGCA-ATCCAGTGTGTGCTGGGCTTTACAGGAAAGAGCT 731
DB 98559 CTGTGTGGCTGAGAGGGCCCGGCGAGATCCAGTGTGTGCTGGGCTTTACAGGAAAGAGCT 98618
QY 732 CCACCTTCT-TGGAGTGTGCAGATGCCATCTAGTGTGTCCACCCGATGGAGCTGGGG 790
DB 98619 CCACCTTCTTGGAGTGTGCAGATGCCATCTAGTGTGTCCACCCGATGGAGCTGGGG 98678
QY 791 CCGGGCAGATGCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCTCTCGACTTCCCTC 850
DB 98679 CCGGGCAGATGCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCTCTCGACTTCCCTC 98738
QY 851 ATTCTCTGCTTGCATCTCCAGCAAGTCTCTCCGTTGGCTGGCTTCATCCACTCTCTC 910
DB 98739 ATTCTCTGCTTGCATCTCCAGCAAGTCTCTCCGTTGGCTGGCTTCATCCACTCTCTC 98798
QY 911 ACTTCTCTGCTTGCATCTCCAGCAAGTCTCTCCGTTGGCTGGCTTCATCCACTCTCTC 951
DB 98799 ACTTCTCTGCTTGCATCTCCAGCAAGTCTCTCCGTTGGCTGGCTTCATCCACTCTCTC 98839

```

```

RESULT 2
AC135044
LOCUS
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
ACCESSION AC135044
VERSION AC135044.1 GI:23505535
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 195476)
  DOE Joint Genome Institute.
  Sequencing of Human Chromosome 16
  Unpublished
  2 (bases 1 to 195476)
  DOE Joint Genome Institute.
  Direct Submission.
  Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
  Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  -----Genome Center
  Center: Joint Genome Institute
  Center Code: JGI
  Web site: http://www.jgi.doe.gov
  -----
Project Information
Center Project Name: 809609
Center clone name: CTD-E1_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 186095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Estimated insert size: 193376; sum-of-contigs estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 22 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 1202: contig of 1202 bp in length
  * 1203 1302: gap of unknown length
  * 1303 2468: contig of 1166 bp in length
  * 2469 2568: gap of unknown length
  * 2569 4077: contig of 1509 bp in length

```





NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCL-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

#### VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-814D15. Actual start of this clone is at base position 1 of RP5-1142J19; actual end is at base position 96975 of RP5-1142J19.

A transposon was identified in the cloning vector for RP5-1142J19.

#### FEATURES

source	Location/Qualifiers
repeat_region	1..96975
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosomes="7"
repeat_region	/map="7q35-q36"
repeat_region	/clone="RP5-1142J19"
repeat_region	/clone_lib="RPCL-5"
repeat_region	532..843
repeat_region	/rpt_family="Alu"
repeat_region	853..883
repeat_region	/rpt_family="AT-rich"
repeat_region	886..923
repeat_region	/rpt_family="Alu"
repeat_region	930..1194
repeat_region	/rpt_family="Alu"
repeat_region	2043..2080
repeat_region	/rpt_family="L1"
repeat_region	2081..2382
repeat_region	/rpt_family="Alu"
repeat_region	2383..2466
repeat_region	/rpt_family="L1"
repeat_region	2467..2780
repeat_region	/rpt_family="Alu"
repeat_region	2781..2796
repeat_region	/rpt_family="L1"
repeat_region	2882..2910
repeat_region	/rpt_family="(CAAAA)n"
repeat_region	2911..3190
repeat_region	/rpt_family="Alu"
repeat_region	3325..3883
repeat_region	/rpt_family="Retroviral"
repeat_region	3926..4113
repeat_region	/rpt_family="MER4-group"
repeat_region	4154..4303
repeat_region	/rpt_family="Alu"
repeat_region	4343..4507
repeat_region	/rpt_family="MER4-group"
repeat_region	4508..4805
repeat_region	/rpt_family="Alu"
repeat_region	4806..5131
repeat_region	/rpt_family="MER4-group"
repeat_region	5132..5445
repeat_region	/rpt_family="Alu"
repeat_region	5446..5466
repeat_region	/rpt_family="MER4-group"
repeat_region	5467..5615
repeat_region	/rpt_family="Retroviral"
repeat_region	6427..6706
repeat_region	/rpt_family="Alu"
repeat_region	6997..7149
repeat_region	/rpt_family="Alu"
repeat_region	7204..7414
repeat_region	/rpt_family="Alu"
repeat_region	7455..7757
repeat_region	/rpt_family="Alu"
repeat_region	7764..7879
repeat_region	/rpt_family="Alu"
repeat_region	7896..7918
repeat_region	/rpt_family="polypurine"
repeat_region	7919..8193
repeat_region	/rpt_family="Alu"
repeat_region	8194..8247
repeat_region	/rpt_family="polypurine"
repeat_region	8339..8594
repeat_region	/rpt_family="Alu"
repeat_region	8682..8966
repeat_region	/rpt_family="Alu"
repeat_region	10832..10737
repeat_region	/rpt_family="MER21_g"
repeat_region	10760..10947
repeat_region	/rpt_family="MER4-group"
repeat_region	10948..11250
repeat_region	/rpt_family="Alu"
repeat_region	11251..11320
repeat_region	/rpt_family="MER4-group"
repeat_region	11321..11487
repeat_region	/rpt_family="Alu"
repeat_region	11488..11785
repeat_region	/rpt_family="Alu"
repeat_region	11786..11905
repeat_region	/rpt_family="Alu"
repeat_region	11906..12285
repeat_region	/rpt_family="MER4-group"
repeat_region	12667..12736
repeat_region	/rpt_family="Alu"
repeat_region	12842..12928
repeat_region	/rpt_family="MER2_type"
repeat_region	13110..13440
repeat_region	/rpt_family="MER4-group"
repeat_region	13447..13486
repeat_region	/rpt_family="(CA)n"
repeat_region	13487..13763
repeat_region	/rpt_family="Alu"
repeat_region	13775..13951
repeat_region	/rpt_family="Alu"
repeat_region	13952..14261
repeat_region	/rpt_family="Alu"
repeat_region	14278..14434
repeat_region	/rpt_family="MER4-group"
repeat_region	14435..14746
repeat_region	/rpt_family="Alu"
repeat_region	14747..14904
repeat_region	/rpt_family="MER4-group"
repeat_region	14932..15070

```

repeat_region      8.0%; Score 79; DB 9; Length 96975;
Best Local Similarity 100.0%; Pred. No. 1.9e-32;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/rpt_family="WER4-group"
15076 .15380
/rpt_family="Alu"

Query Match
Best Local Similarity 100.0%; Pred. No. 1.9e-32;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CGATCTCTGACCTCGGATCGCCGCCGCTCAGCTCCCAAGTGTGGATTACAGGCA 93
Db 16583 CGATCTCTGACCTCGGATCGCCGCCGCTCAGCTCCCAAGTGTGGATTACAGGCA 16524

QY 94 TGAGCCACCGCGCCGCCGCC 112
Db 16523 TGAGCCACCGCGCCGCCGCC 16505

RESULT 4
AC118269 129758 bp DNA linear PRI 24-JUN-2002
LOCUS Homo sapiens chromosome 17, clone CTD-2383D6, complete sequence.
DEFINITION AC118269
ACCESSION AC118269
VERSION AC118269.6 GI:21536064
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129758)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone CTD-2383D6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 129758)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 129758)
TITLE RepeatMasker
JOURNAL
REFERENCE 1 (bases 1 to 129758)
AUTHORS Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26202
Center clone name: 2383_D_6
----- Location/Qualifiers
1.129758
/clone="CTD-2383D6"
/clone_lib="CITD1 Human BAC"
/complement(2..224)
/rpt_family="AluJo"
/complement(365..413)
/rpt_family="L2"
repeat_region 415..546
/rpt_family="AluSg/x"
repeat_region 549..844
/rpt_family="AluSp"
repeat_region 851..1158
/rpt_family="AluX"
repeat_region complement(1177..1928)
/rpt_family="L2"
repeat_region 5198..5315
/rpt_family="FLAM_A"

```

```

O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 129758)
TITLE RepeatMasker
JOURNAL
REFERENCE 1 (bases 1 to 129758)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 21, 2002 this sequence version replaced gi:21392520.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26202
Center clone name: 2383_D_6
----- Location/Qualifiers
1.129758
/clone="CTD-2383D6"
/clone_lib="CITD1 Human BAC"
/complement(2..224)
/rpt_family="AluJo"
/complement(365..413)
/rpt_family="L2"
repeat_region 415..546
/rpt_family="AluSg/x"
repeat_region 549..844
/rpt_family="AluSp"
repeat_region 851..1158
/rpt_family="AluX"
repeat_region complement(1177..1928)
/rpt_family="L2"
repeat_region 5198..5315
/rpt_family="FLAM_A"

```

```

repeat_region complement(5370..5680)
unsure /rpt_family="AluSg"
5466..5484 /note="single clone coverage"
repeat_region complement(6042..6349)
repeat_region /rpt_family="AluY"
repeat_region complement(6856..7014)
repeat_region /rpt_family="AluJb"
repeat_region complement(7020..7326)
repeat_region /rpt_family="AluSx"
7404..7719 /rpt_family="AluSx"
repeat_region /rpt_family="AluSx"
repeat_region complement(8263..8548)
repeat_region /rpt_family="AluSx"
repeat_region complement(8550..8760)
repeat_region /rpt_family="L1PAL0"
8828..9158 /rpt_family="AluSg"
repeat_region /rpt_family="AluSg"
repeat_region complement(9234..9436)
repeat_region /rpt_family="MER20"
repeat_region complement(10017..10326)
repeat_region /rpt_family="AluSx"
10476..10781 /rpt_family="AluSx"
repeat_region /rpt_family="AluSx"
repeat_region complement(11225..11484)
repeat_region /rpt_family="L2"
repeat_region complement(11549..11657)
repeat_region /rpt_family="MER105"
repeat_region complement(11658..11969)
repeat_region /rpt_family="AluSg"
repeat_region complement(11970..12000)
repeat_region /rpt_family="MER105"
repeat_region complement(12001..12099)
repeat_region /rpt_family="LTR65"
repeat_region complement(12100..12401)
repeat_region /rpt_family="AluSx"
repeat_region complement(12402..12750)
repeat_region /rpt_family="LTR65"
repeat_region complement(12751..13057)
repeat_region /rpt_family="AluY"
repeat_region complement(13058..13208)
repeat_region /rpt_family="LTR65"
repeat_region complement(13213..13285)
repeat_region /rpt_family="MER105"
repeat_region complement(13708..13847)
repeat_region /rpt_family="AluJb"
repeat_region complement(13848..14151)
repeat_region /rpt_family="AluSp"
repeat_region complement(14152..14294)
repeat_region /rpt_family="AluJb"
repeat_region complement(14427..14567)
repeat_region /rpt_family="MIR"
14568..14845 /rpt_family="AluSx"
repeat_region /rpt_family="AluSx"
14846..15018 /rpt_family="AluSg"
repeat_region /rpt_family="L3"
repeat_region complement(15478..15525)
repeat_region /rpt_family="AluY"
repeat_region complement(15588..15696)
repeat_region /rpt_family="L2"
15697..16005 /rpt_family="AluSx"
repeat_region /rpt_family="AluSx"
repeat_region complement(16006..16315)
repeat_region /rpt_family="L2"
repeat_region complement(16879..17153)
repeat_region /rpt_family="AluSg"
17227..17421 /rpt_family="MER3"
repeat_region /rpt_family="MER3"
17577..18116 /rpt_family="L2"
repeat_region complement(19052..19360)

```

Query Match 6.7%; Score 66; DB 9; Length 129758;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-25;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CGATCTCTGACCTCGTATCCGCCCGCTCAGCTTCCCAAGTCTGGGATTACAGGCA 93  
 Db 95562 CGATCTCTGACCTCGTATCCGCCCGCTCAGCTTCCCAAGTCTGGGATTACAGGCA 93

QY 94 TGAGCC 99  
 Db 95502 TGAGCC 95497

RESULT 5  
 AC116914/c 138536 bp DNA linear PRI 06-AUG-2002  
 LOCUS Homo sapiens chromosome 17, clone CTD-319515, complete sequence.  
 DEFINITION AC116914  
 ACCESSION AC116914  
 VERSION AC116914.5 GI:22038652  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (Bases 1 to 138536)  
 AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 17, clone CTD-319515  
 JOURNAL Unpublished  
 REFERENCE 2 (Bases 1 to 138536)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Bouckhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Gagegan,J.J., Gardyna,S.,  
 Ginde,S., Gordon,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
 Minova,I., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (Bases 1 to 138536)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Bouckhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fero,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Gordon,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliiev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,  
 Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,  
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,  
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,

Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

4 (bases 1 to 138536)

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

Submitted (01-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

5 (bases 1 to 138536)

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Aug 1, 2002 this sequence version replaced gi:21591993.  
All repeats were identified using RepeatMasker:  
Snit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L26214

Center clone name: 3195\_I\_5

## FEATURES

----- Location/Qualifiers

1. .138536

/organism="Homo sapiens"

/mol type="genomic DNA"

/db xref="taxon:9606"

/chromosome="17"

/map="17"

/clone="CTD-3195I5"

/clone.lib="CITD2 Human BAC"  
complement(240..399)  
/rpt family="AluDb"  
complement(400..691)  
/rpt family="AluSc"  
complement(692..833)  
/rpt family="AluDb"  
complement(1109..1343)  
/rpt family="AluSc"  
1344..1572  
/rpt family="Charlie4a"  
2215..2517  
/rpt family="AluY"  
complement(3386..3656)  
/rpt family="AluSc"  
3941..4060  
/rpt family="FLAM\_A"  
4068..4385  
/rpt family="AluSg"  
4583..4895  
/rpt family="AluSc"  
5084..5389  
/rpt family="AluSc"  
complement(5809..6121)  
/rpt family="AluSc"  
complement(6125..6286)  
/rpt family="AluSg/x"  
6975..7099  
/rpt family="L1M4"  
7100..7393  
/rpt family="AluSc"  
7394..7519  
/rpt family="L1M4"  
7520..7817  
/rpt family="AluJo"  
7818..7907  
/rpt family="L1M4"  
7908..8216  
/rpt family="AluSp"  
8217..8311  
/rpt family="L1M4"  
complement(8377..8664)  
/rpt family="AluY"  
complement(8667..8963)  
/rpt family="AluSc"  
10023..10292  
/rpt family="AluJo"  
10313..10629  
/rpt family="AluSg"  
12082..12156  
/rpt family="L13"  
12589..12849  
/rpt family="MLTIL"  
13297..13387  
/rpt family="L1MB4A"  
13388..13703  
/rpt family="AluSc"  
complement(14154..14464)  
/rpt family="AluSc"  
complement(15023..15320)  
/rpt family="AluSc"  
complement(15377..15666)  
/rpt family="AluJo"  
complement(15699..16011)  
/rpt family="AluJo"  
complement(16152..16196)  
/rpt family="L2"  
16202..16333  
/rpt family="AluSg/x"  
16336..16631

Query Match

Best Local Similarity 100.0%; Score 66; DB 9; Length 138536;  
Pred. No. 3.5e-25;



Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zairoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 4, 2001 this sequence version replaced gl:13959235.  
All repeats were identified using RepeatMasker:  
Smit, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project information

Center project name: L487  
Center clone name: 231\_C16

----- Summary Statistics

Sequencing vector: M13; M77815; 41% of reads  
Sequencing vector: Plasmid; L08752; 4% of reads  
Sequencing vector: Plasmid; n/a; 55% of reads  
Chemistry: Dye-primer-amersham; 8% of reads  
Chemistry: Dye-terminator Big Dye; 92% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 203261 bases at least Q40  
Consensus quality: 205916 bases at least Q30  
Consensus quality: 207122 bases at least Q20  
Insert size: 177000; agarose-fp  
Insert size: 208485; su

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 2927: contig of 2927 bp in length
2 3027: gap of 100 bp
3 3028 4189: contig of 1162 bp in length
4 4190 4289: gap of 100 bp
5 4290 5459: contig of 1180 bp in length
6 5470 5589: gap of 100 bp
7 5590 6656: contig of 1087 bp in length
8 6657 6756: gap of 100 bp
9 6757 8150: contig of 1394 bp in length
10 8151 8250: gap of 100 bp
11 8251 9645: contig of 1385 bp in length
12 9646 9746: gap of 100 bp
13 9746 11119: contig of 1374 bp in length
14 11120 11219: gap of 100 bp
15 11220 13069: contig of 1850 bp in length
16 13070 13169: gap of 100 bp
17 13170 15065: contig of 1896 bp in length
18 15066 15165: gap of 100 bp
19 15166 20167: contig of 5002 bp in length
20 20168 20267: gap of 100 bp
21 20268 29608: contig of 9341 bp in length
22 29609 29708: gap of 100 bp
23 29709 45638: contig of 15930 bp in length
24 45639 45738: gap of 100 bp
25 45739 73550: contig of 27812 bp in length
26 73551 73650: gap of 100 bp
27 73651 138571: contig of 64921 bp in length
28 138572 138671: gap of 100 bp
29 138672 209885: contig of 71214 bp in length.

```

## FEATURES source

```

1. .209885
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="17"
   /map="17"

```

/clone="RP11-231C16"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .2927

## misc\_feature

/note="assembly\_fragment"  
clone end:SP6  
vector side:left

## misc\_feature

3028. .4189  
/note="assembly\_fragment"

## misc\_feature

4290. .5469  
/note="assembly\_fragment"

## misc\_feature

5570. .6656  
/note="assembly\_fragment"

## misc\_feature

6757. .8150  
/note="assembly\_fragment"

## misc\_feature

8251. .9645  
/note="assembly\_fragment"

## misc\_feature

9746. .11119  
/note="assembly\_fragment"

## misc\_feature

11220. .13069  
/note="assembly\_fragment"

## misc\_feature

13170. .15065  
/note="assembly\_fragment"

## misc\_feature

15166. .20167  
/note="assembly\_fragment"

## misc\_feature

20268. .29608  
/note="assembly\_fragment"

## misc\_feature

29709. .45638  
/note="assembly\_fragment"

## misc\_feature

45739. .73550  
/note="assembly\_fragment"

## misc\_feature

73651. .138571  
/note="assembly\_fragment"

## misc\_feature

138672. .209885  
/note="assembly\_fragment"

## ORIGIN

Query Match 6.7%; Score 66; DB 2; Length 209885;  
Best Local Similarity 100.0%; Pred. No. 3.6e-25;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 CGATCTCTGACCTCGATCCGCCGCCCTCCAGCTTCCCAAGCTGGGATTACAGCA 93

Db 181394 CGATCTCTGACCTCGATCCGCCGCCCTCCAGCTTCCCAAGCTGGGATTACAGCA 181453

Oy 94 TGAGCC 99

Db 181454 TGAGCC 181459

## RESULT 8

HS20208

LOCUS

DEFINITION

Human DNA sequence from clone RP1-20208 on chromosome 1p36.11-36.31  
Contains the 5' part of a gene for a novel rat Espin Like protein  
containing Ankrepeats, the gene for the ortholog of rodent HES2  
(Hairy and Enhancer of Split 2) and the 5' end of the gene for HBACH  
(Brain Acyl-CoA Hydrolase (Acyl Coenzyme A Thioester Hydrolase, EC  
3.1.2.2). Contains ESTs, GSSs and putative CpG islands, complete  
sequence.

ACCESSION

AL031848

VERSION

AL031848.11 GI:4914512

KEYWORDS

HTG; Acyl Coenzyme A Thioester Hydrolase; Ankrepeat; Brain  
Acyl-CoA Hydrolase; Enhancer of Split; Espin; Hairy; HBACH; HES2.

SOURCE

Homo sapiens

ORGANISM

1 (bases 1 to 143065)

REFERENCE

Howden, P.

AUTHORS

Direct Submission

TITLE

Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,

JOURNAL

Cambridge, CB10 1SA, UK. E-mail enquiries:

[humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

## COMMENT

On May 28, 1999 this sequence version replaced gi:4678811.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>

RP1-20208 is from the library RPCT-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pCVPAC2

This sequence is the entire insert of clone RP1-20208 The true left end of clone RP1-120232 is at 135606 in this sequence.

## FEATURES

source  
1..143065  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="RZPD:RPCTP704008202"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.11-36.31"  
/clone="RP1-20208"  
/clone\_lib="RPCT-1"  
7..125  
/note="MIR repeat: matches 13. .144 of consensus"  
repeat\_region  
127..219  
/note="MIR repeat: matches 14. .106 of consensus"  
repeat\_region  
365..404  
/note="Alu repeat: matches 126. .165 of consensus"  
repeat\_region  
405..634  
/note="AluSg/x repeat: matches 83. .309 of consensus"  
misc\_feature  
complement(635..907)  
/note="match: GSS: Em:B42699"  
misc\_feature  
complement(635..895)  
/note="match: GSS: Em:B47252"  
gene  
complement(732..27887)  
/genes="dJ20208.1"  
mRNA  
complement(join<732..995,1080..1225,3737..3935,4026..4187,6893..7164,8146..8347,11763..11894,12020..12202,12388..12574,24409..24602,27579..27887))  
/genes="dJ20208.1"  
/products="dJ20208.1 (novel rat Espin LIKE protein containing Ank repeats)"  
/note="match: CDSAs: Em:U46007 Em:AL035288 Em:AF076856 Em:AF134401  
match: ESTs: Em:H72801"  
/evidence="not\_experimental"  
complement(join<732..995,1080..1225,3737..3935,4026..4187,6893..7164,8146..8347,11763..11894,12020..12202,12388..12574,24409..24602,27579..27872))

/gene="dJ20208.1"  
/note="match: proteins: Tr:Q63618 Tr:Q24341 Tr:O89019 Sw:Q01484 Tr:Q17343 Sw:Q01485 Tr:Q25338 Tr:Q92625"  
/codon\_start=1  
/evidence="not\_experimental"  
/product="dJ20208.1 (novel rat Espin LIKE protein containing Ank repeats)"  
/protein\_id="CAB46197.1"  
/db\_xref="GI:5327035"  
/translations="MALEQALQARQGLDVLRLHAAGLGLPSLRDLPLDALPVHAA  
RAGKLCRLVEALPAARARNATPAHDASATGHLACLQWLQSGGCRVQDKDN  
SGATVLAARFGHPEVNVNHHGGDPTAATDMGALPIHYAAAGDPFSLKLLVHE  
YFEGVNAQTNGATPLYLACQEGHLEVTVLQVECGADPHARADQMTFLHAAQGH  
SPVIVLVSTQDVSQKDXGATAMHFAASRGHTKVLWLLHGGELISADLWGTPL  
HDAANGELCCQILVNVGAELDRDRDGTAAADLDFNGHSHCTYLRVTENLSVEH  
RVLSRDPSEALEAKQPDGMSSTTVVQPLNFDLSSPTSLSNVDSGSSHSIKG  
CHPCGLSSARAADIOSYMDMLNPELGLPRGTIGKTPPPPPPPPPPPPTQIAPP  
PFGYAPAPFPVQADIIYMTKRLHVETALKEKLSGCDHGLRQDSSRKPR  
AFKQSTGYRQLGRCPGCTLAARPGMAHSEEAALLPNNHVPNCGAADPRAARLP  
PPPPPPPPPEAAASPPPPAPLPPLPESAGCGQRRSSSTGKTFNMSPTGDNSE  
LLAEIRAGSLKPTQSGKLTTFVSGIQGPAQPDSPSPSPSPSPSPSPSPSPSP  
QPLILGSLVPEVPTTPAPGVQLDVEALIPTHDEQGRPIPEWKQVNVKMLQVQEE  
EQRK"  
778..1198  
/note="CpG island"  
/evidence="not\_experimental"  
2896..3126  
/note="match: GSS: Em:AQ036085"  
3347..3415  
/note="3 copies 23 mer 81% conserved"  
3441..3507  
/note="MIR repeat: matches 57. .121 of consensus"  
3702..5680  
/note="CpG island"  
/evidence="not\_experimental"  
5573..5686  
/note="6 copies 19 mer 91% conserved"  
6023..6423  
/note="L1M3e repeat: matches -370. .12 of consensus"  
6999..7074  
/note="38 copies 2 mer gg 71% conserved"  
7708..7741  
/note="17 copies 2 mer gt 82% conserved"  
7709..7888  
/note="4 copies 45 mer 68% conserved"  
9466..9591  
/note="MIR repeat: matches 31. .162 of consensus"  
9640..9931  
/note="AluSg repeat: matches 1. .299 of consensus"  
9940..10240  
/note="AluSg repeat: matches 1. .300 of consensus"  
10241..10539  
/note="AluSg repeat: matches 1. .294 of consensus"  
10556..10642  
/note="MIR repeat: matches 173. .261 of consensus"  
11758..12719  
/note="CpG island"  
/evidence="not\_experimental"  
12259..12372  
/note="57 copies 2 mer cc 66% conserved"  
12810..13093  
/note="AluSg repeat: matches 3. .290 of consensus"  
13220..13427  
/note="L2 repeat: matches 2414. .2615 of consensus"  
13434..13624  
/note="MER53 repeat: matches 1. .189 of consensus"  
13693..13998  
/note="AluSx repeat: matches 1. .306 of consensus"  
14150..14201  
/note="MIR repeat: matches 68. .125 of consensus"  
14376..14517  
/note="MIR repeat: matches 100. .244 of consensus"  
15679..15715



```

repeat_region /note="MIR repeat: matches 218. .254 of consensus"
16039. .16334
repeat_region /note="AluJo repeat: matches 18. .311 of consensus"
16409. .16636
repeat_region /note="MIR repeat: matches 32. .261 of consensus"
16832. .17009
repeat_region /note="MIR repeat: matches 51. .244 of consensus"
17818. .18085
repeat_region /note="L2 repeat: matches 2311. .2605 of consensus"
18626. .18913
repeat_region /note="AluSx repeat: matches 1. .291 of consensus"
18925. .19128
repeat_region /note="L2 repeat: matches 2324. .2555 of consensus"
19202. .19370
repeat_region /note="AluSg/x repeat: matches 134. .302 of consensus"
19372. .19665
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
19679. .19986
repeat_region /note="AluSg repeat: matches 2. .308 of consensus"
complement(20236. .20676)
misc_feature /genes="dJ20208.1"
/note="match: GSS: Em:B43196"
complement(20501. .20541)
misc_feature /genes="dJ20208.1"
/note="match: GSS: Em:B43451"
20905. .21014
repeat_region /note="L2 repeat: matches 2593. .2706 of consensus"
21016. .21075
repeat_region /note="MIR repeat: matches 130. .191 of consensus"
21171. .21501
repeat_region /note="AluSx repeat: matches 1. .311 of consensus"
22030. .22214

Query Match 6.4%; Score 63; DB 9; Length 143065;
Best Local Similarity 100.0%; Pred. No. 1.7e-23;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 TCTGACCTCGTATCGCCGCGCTCAGCTTCCAAAGCTCGGATTACAGCATGAC 98
Db 19592 TCTGACCTCGTATCGCCGCGCTCAGCTTCCAAAGCTCGGATTACAGCATGAC 19651

QY 99 CAC 101
Db 19652 CAC 19654

RESULT 9
HS57G9/c
LOCUS
DEFINITION Human DNA sequence from clone CTA-57G9 on chromosome 22q12.1,
complete sequence.
ACCESSION Z95116
VERSION Z95116.1 GI:2832592
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113872)
McLaren,S.
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 5, 1998 this sequence version replaced gi:2578126.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

```

```

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-57G9 is from
the human BAC library described in U-J. Kim et al. (1996) Genomics
34, 213-218.
VECTOR: pBAC108L
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

This sequence is the entire insert of clone CTA-57G9. The true right end of clone RPS-117M9 is at 100 in this sequence.

#### FEATURES

```

Source
1. 113872
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-57G9"
/clone_lib="CIT978SK-A1"
17. .392
/note="match: GSS: Em:B14225"
repeat_region /note="2.0 copies 6 mer GCCATG 24% conserved"
32. .43
repeat_region /note="3.3 copies 3 mer ATC 20% conserved"
171. .180
complement(305. .682)
misc_feature /note="match: GSS: Em:B14096"
complement(387. .678)
misc_feature /note="match: GSS: Em:B14098"
complement(430. .682)
misc_feature /note="match: GSS: Em:B36168"
593. .618
repeat_region /note="13.0 copies 2 mer GT 43% conserved"
740. .1035
repeat_region /note="Alu repeat: matches 1. .296 of consensus"
1177. .1188
repeat_region /note="2.0 copies 6 mer GAAAGG 24% conserved"
complement(1326. .1866)
repeat_region /note="MLN1F2 repeat: matches 33. .554 of consensus"
complement(1867. .2202)
repeat_region /note="MLN1A repeat: matches 24. .374 of consensus"
complement(2221. .2519)
repeat_region /note="WER21C repeat: matches 666. .932 of consensus"
2520. .2659
repeat_region /note="FLAM C repeat: matches 1. .129 of consensus"
complement(2660. .3073)
repeat_region /note="WER21C repeat: matches 21. .666 of consensus"
3241. .3452
repeat_region /note="MLN1D repeat: matches 1. .236 of consensus"
3461. .4005
repeat_region /note="13.6 copies 40 mer
AGAGTGATGATGAAGGAAACAGGAGGAGAGGTGATGA 780% conserved"
4013. .4345
repeat_region /note="MLN1D repeat: matches 180. .505 of consensus"
4553. .4566
repeat_region /note="14.0 copies 1 mer T 28% conserved"
4615. .4629

```

```

/notes="3.0 copies 5 mer CAAA 21% conserved"
join(4811..4935,8717..8870,20796..21128,22078..22209)
/genes="BX5769.1"
CDS
join(4811..4935,8717..8870,20796..21128,22078..22209)
/genes="BX5769.1"
/notes="supported by GENSCAN"
match: ESTs: Em:AA506741 Em:AA670682
match: proteins: Tr:P91972 Sw:P98085 Tr:Q15113 Tr:Q35113
Tr:Q24132 Tr:O70244 Tr:O57433 Tr:O88204 Sw:P97435
Tr:O60494 Tr:O57658 Tr:O57074 Tr:O57434 Tr:O08628
Tr:Q61398 Sw:P98066 Tr:Q23995 Sw:P97333 Tr:O08859
Tr:O95917"
/codon_start=3
/evidence=not experimental
/product="BX5769.1 (novel Kringle and CUB domain protein)"
/protein_id="CAB62952.1"
/db_xref="GI:6572252"
/translation="PGLGICYDHGPNPPLTGTSTKSLTIQTICISFCRSQRPKFAQ
MESGYACFCGNPDYKWKYGAASCTCNVCFGDHTQPCGGDRIILFTLVGACGCGNY
SAMSVYSPDPTVATGRCVCTIRVPFGASHIFSPPLFDIRDSADWVLLDGYTH
EVLARFHGSRPLSNVSLDFVLLYFSPDRINQAGFAVLQAVKEELPQERPAVQ
TVAEVLTEQANUSVSARSKVLVIITS"
4995..5311
/notes="LIMB7 repeat: matches 5850..6194 of consensus"
5333..6111
/notes="LIME1 repeat: matches 5043..5827 of consensus"
6112..6420
/notes="ALUSg repeat: matches 1..306 of consensus"
6421..6501
/notes="LIME1 repeat: matches 5827..5906 of consensus"
6502..6965
/notes="LIME1 repeat: matches 4987..5457 of consensus"
6966..7272
/notes="ALUSx repeat: matches 3..311 of consensus"
7273..7946
/notes="LIME1 repeat: matches 5467..6068 of consensus"
complement(7947..8246)
/notes="ALUSx repeat: matches 1..300 of consensus"
8247..8291
/notes="LIME1 repeat: matches 6068..6113 of consensus"
8327..8426
/notes="MIR repeat: matches 134..262 of consensus"
8941..8952
/notes="3.0 copies 4 mer CCT 24% conserved"
8960..8972
/notes="3.2 copies 4 mer ATTC 26% conserved"
complement(8997..9100)
/notes="L2 repeat: matches 3138..3259 of consensus"
9244..9253
/notes="10.0 copies 1 mer T 20% conserved"
9570..9582
/notes="13.0 copies 1 mer A 26% conserved"
complement(9708..10119)
/notes="MLT1C repeat: matches 8..465 of consensus"
complement(10556..10724)
/notes="MIR repeat: matches 74..252 of consensus"
10841..10916
/notes="LIMB7 repeat: matches 5508..5583 of consensus"
10917..11122
/notes="ALUSg/x repeat: matches 83..296 of consensus"
11123..11661
/notes="LIMB7 repeat: matches 5583..6122 of consensus"
11710..11811
/notes="Tiggers repeat: matches 2307..2406 of consensus"
11816..11834
/notes="2.1 copies 9 mer TTTAAATGG 38% conserved"
complement(12111..12286)
/notes="MIR repeat: matches 47..262 of consensus"
12663..12810
/notes="ALUSg/x repeat: matches 139..230 of consensus"
complement(13522..13690)
/notes="MIR repeat: matches 78..252 of consensus"
complement(13684..13732)

```

```

/notes="L2 repeat: matches 3182..3229 of consensus"
complement(13805..13964)
/notes="MER104 repeat: matches 1..176 of consensus"
14449..14464
/notes="2.3 copies 7 mer GTTAAAG 23% conserved"
complement(14535..14785)
/notes="MIR repeat: matches 5..259 of consensus"
complement(14788..14982)
/notes="LIME1 repeat: matches 6741..6952 of consensus"
14983..15002
/notes="10.0 copies 2 mer TA 40% conserved"
15003..15057
/notes="LTR39-int repeat: matches 3905..3959 of consensus"
15067..15093
/notes="13.5 copies 2 mer TA 36% conserved"
15069..15091
/notes="2.9 copies 8 mer TATATATG 46% conserved"
15100..15189
/notes="LIMC4 repeat: matches 6614..6703 of consensus"
15158..15230
/notes="LIMCa repeat: matches 1995..2064 of consensus"
complement(15233..15319)
/notes="LIME1 repeat: matches 6659..6744 of consensus"
complement(15320..15625)
/notes="ALUSp repeat: matches 1..307 of consensus"
complement(15626..16420)
/notes="LIME1 repeat: matches 5869..6659 of consensus"
16421..16729
/notes="ALUy repeat: matches 1..309 of consensus"

Query Match 6.1%; Score 60; DB 9; Length 113872;
Best Local Similarity 100.0%; Pred. No. 7.8e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GTGATCCGCGCGCTCCCAAGTCTGGATTACAGCATGAGCCACCGGCGCC 108
|||||
DB 112880 GTGATCCGCGCGCTCCCAAGTCTGGATTACAGCATGAGCCACCGGCGCC 112821

RESULT 10
AC004645/c
LOCUS Homo sapiens chromosome 16, cosmid clone 400C4 (LANL), complete
DEFINITION Homo sapiens chromosome 16, cosmid clone 400C4 (LANL), complete
sequence.
ACCESSION AC004645
VERSION AC004645.1 GI:3097839
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 44174)
Ricke,D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
2 (bases 1 to 44174)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Misra,M. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
Unpublished
3 (bases 1 to 44174)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Misra,M. and Deaven,L.
Direct Submission
Submitted (01-MAY-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los

```

```

FEATURES             Alamos, NM 87545, USA
    source            Location/Qualifiers
1..44174
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="16"
    /map="16p13.3"
    /clone="400C4"
    complement(355..1214)
    /standard_name="fmc17"
    /note="96% identity dbEST:AA631950"
416..458
    /note="GRAIL 2 excellent exon, frame 2"
962..1020
    /note="GRAIL 2 excellent exon, frame 0"
1396..1508
    /note="87% identity no80a12.s1"
    /db_xref="dbEST:AA604650"
    complement(1867..2144)
    /note="87% identity X59618"
    complement(3192..3576)
    /rpt_family="Alu"
    complement(3618..4137)
    /rpt_family="Alu"
4278..4599
    /rpt_family="Alu"
4569..4588
    /note="(A)20"
    /rpt_type=tandem
    /rpt_unit="A"
5384..5837
    /note="99% identity EST z131a04.s1"
    /db_xref="dbEST:AA706919"
    complement(6884..7023)
    /rpt_family="Alu"
    complement(7093..7725)
    /rpt_family="Alu"
7811..8117
    /rpt_family="Alu"
9238..9330
    /note="GRAIL 2 excellent exon, frame 1"
    complement(10614..10716)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(11928..12071)
    /rpt_family="Alu"
13214..13542
    /rpt_family="Alu"
13711..13750
    /note="(AC)20"
    /rpt_type=tandem
    /rpt_unit="AC"
14070..14157
    /rpt_family="L1MCB"
14467..14709
    /rpt_family="L1M4"
    complement(14704..15013)
    /rpt_family="Alu"
    complement(15018..15315)
    /rpt_family="Alu"
15367..15604
    /rpt_family="L1MCB"
    complement(16360..16708)
    /rpt_family="Alu"
16375..16397
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
16724..16794
    /rpt_family="Alu"
16784..16996
    /rpt_family="7SL"
16978..17327
    /rpt_family="Alu"

repeat_region       /rpt_family="Alu"
17328..17635
    /rpt_family="Alu"
17745..18053
    /rpt_family="Alu"
    complement(18227..18533)
    /rpt_family="Alu"
18809..18917
    /rpt_family="L1M2"
18954..19282
    /rpt_family="Alu"
19278..19798
    /rpt_family="L1ME"
19948..20656
    /rpt_family="Alu"
20650..20833
    /rpt_family="Alu"
    complement(20981..21127)
    /rpt_family="Alu"
21330..21684
    /rpt_family="Alu"
21714..22048
    /rpt_family="Alu"
22060..22632
    /rpt_family="L1"
    complement(22661..22976)
    /rpt_family="Alu"
22661..22683
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
22994..23307
    /rpt_family="Alu"
23540..24242
    /rpt_family="L1"
23894..23916
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
24256..24578
    /rpt_family="Alu"
24635..24987
    /rpt_family="Alu"
    complement(25737..26036)
    /rpt_family="Alu"
    complement(26209..26434)
    /rpt_family="Alu"
    complement(26443..26829)
    /rpt_family="Alu"
26859..27193
    /rpt_family="Alu"
    complement(27569..27877)
    /rpt_family="Alu"
    complement(27852..28119)
    /rpt_family="L1"
    complement(28107..28242)
    /rpt_family="Alu"
    complement(28199..28953)
    /rpt_family="L1"
29073..29237
    /rpt_family="Alu"
    complement(29616..29932)
    /rpt_family="Alu"
30472..31073
    /rpt_family="Alu"
    complement(31309..31412)
    /rpt_family="Alu"
31636..31943
    /rpt_family="Alu"
32031..32566
    /rpt_family="Alu"
32519..33006
    /rpt_family="Alu"

```



```

misc_feature      3454..3585
/note="match: GSS: Em:AL276849"
repeat_region    3490..3581
/note="23 copies 4 mer tggg 84% conserved"
repeat_region    3581..3592
/note="TH1A repeat: matches 1..353 of consensus"
misc_feature      3592..3689
/note="match: GSS: Em:B53127"
/note="complement(4751..5369)"
misc_feature      3689..3790
/note="match: GSS: Em:AQ242433"
/note="complement(5089..5368)"
repeat_region    3790..3889
/note="MER5A repeat: matches 1..186 of consensus"
repeat_region    3889..3988
/note="MIR repeat: matches 46..228 of consensus"
misc_feature      3988..4087
/note="match: STS: Em:HSPF1266"
repeat_region    4087..4186
/note="LIR33 repeat: matches 172..514 of consensus"
misc_feature      4186..4285
/note="match: GSS: Em:AQ535194"
misc_feature      4285..4384
/note="match: GSS: Em:AQ548736"
repeat_region    4384..4483
/note="12 copies 4 mer gaag 85% conserved"
repeat_region    4483..4582
/note="12 copies 4 mer tata 81% conserved"
repeat_region    4582..4681
/note="21 copies 2 mer at 83% conserved"
repeat_region    4681..4780
/note="9 copies 4 mer atat 88% conserved"
repeat_region    4780..4879
/note="MER5A repeat: matches 1..189 of consensus"
repeat_region    4879..4978
/note="MER5A repeat: matches 5..71 of consensus"
repeat_region    4978..5077
/note="MER5A repeat: matches 1..299 of consensus"
misc_feature      5077..5176
/note="match: GSS: Em:AQ147161"
repeat_region    5176..5275
/note="AluX repeat: matches 1..307 of consensus"
repeat_region    5275..5374
/note="MER5A repeat: matches 9..189 of consensus"
repeat_region    5374..5473
/note="LIMC3 repeat: matches 7324..7733 of consensus"
misc_feature      5473..5572
/note="match: GSS: Em:AQ184417"
repeat_region    5572..5671
/note="MIR repeat: matches 31..197 of consensus"
repeat_region    5671..5770
/note="MER5A repeat: matches 64..168 of consensus"
repeat_region    5770..5869
/note="MER5A repeat: matches 1..224 of consensus"
repeat_region    5869..5968
/note="MER5A repeat: matches 55..188 of consensus"
repeat_region    5968..6067
/note="MIR repeat: matches 3..390 of consensus"
misc_feature      6067..6166
/note="match: GSS: Em:AQ390380"
repeat_region    6166..6265
/note="LIME1 repeat: matches 5708..6158 of consensus"
repeat_region    6265..6364
/note="LIME1 repeat: matches 4597..5629 of consensus"
repeat_region    6364..6463
/note="LIME1 repeat: matches 4497..6175 of consensus"
repeat_region    6463..6562
/note="LIME5 repeat: matches 5259..6142 of consensus"
repeat_region    6562..6661
/note="LIPAG repeat: matches 4945..5239 of consensus"
repeat_region    6661..6760
/note="LIPAG repeat: matches 2765..3567 of consensus"
repeat_region    6760..6859
/note="LIPAG repeat: matches 1..295 of consensus"
repeat_region    6859..6958
/note="LIPAG repeat: matches 1..295 of consensus"

```

```

/note="LIM4 repeat: matches 2448..2765 of consensus"
complement(26705..26933)
/note="match: GSS: Em:AQ245856"
27193..27364
/note="LIM4 repeat: matches 2309..2474 of consensus"
27365..27864
/note="LIP2 repeat: matches 5646..6154 of consensus"
27865..27900
/note="LIM4 repeat: matches 2273..2309 of consensus"
27901..28132
/note="58 copies 4 mer atat 59% conserved"
27937..28080
/note="9 copies 16 mer f3% conserved"
complement(28031..28412)
/note="match: GSS: Em:A2001959"

Query Match      6.0%; Score 59; DB 9; Length 45454;
Best Local Similarity 100.0%; Pred.No. 2.7e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34  CGATCTCTGACCTGCTGATCCGCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGC 92
          |||||
Db      43027  CGATCTCTGACCTGCTGATCCGCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGC 42969

RESULT 12
LOCUS    AF212832/c
DEFINITION Homo sapiens chromosome 8 clone CTC-806C5 map 8q24.3, complete
          sequence.
ACCESSION AF212832
VERSION   AF212832.5 GI:25100909
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 122029)
           Blechschmidt,K., Schattevoy,R., Baumgart,C. and Rosenthal,A.
           Unpublished
REFERENCE 2 (bases 1 to 122029)
           Blechschmidt,K., Wen,G., Schilhabel,M., Baumgart,C., Menzel,U.,
           Deter,M. and Rosenthal,A.
           Direct Submission
           Submitted (08-DEC-1999) Genome Analysis, Institute of Molecular
           Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 122029)
           Lagemann,D. and Platzer,M.
           Direct Submission
           Submitted (10-SEP-2002) Genome Analysis, Institute of Molecular
           Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 122029)
           Lagemann,D. and Platzer,M.
           Direct Submission
           Submitted (19-NOV-2002) Genome Analysis, Institute of Molecular
           Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT   On Nov 19, 2002 this sequence version replaced gi:272773267.
           ----- Genome Center
           Center: Institute of Molecular Biotechnology
           Center code: IMB
           Web site: http://genome.imb-jena.de/
           Contact: gscj-submit@genome.imb-jena.de
           ----- Project Information
           Center project name: H283
           Center clone name: CTC-806C5
           ----- Summary Statistics
           Sequencing vector: pUC18; 100% of reads
           Assembly program: Phrap; version 0.990329
           Consensus quality: 121876 bases at least Q40
           Consensus quality: 121937 bases at least Q30
           Consensus quality: 121979 bases at least Q20
           Quality coverage: 10.81x

```

-----  
 This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.  
 -----

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
 -----

## FEATURES

## Location/Qualifiers

```

source      1..122029
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="8"
             /map="8q24.3"
             /clone="CTC-806C5"
variation   844
             /note="T substituted in clone: CTC-806C5"
             /replace="G"
variation   887
             /note="T substituted in clone: CTC-806C5"
             /replace="A"
variation   1172
             /note="T substituted in clone: CTC-806C5"
             /replace="C"
variation   1278
             /note="deleted in clone: CTC-806C5"
             /replace="G"
variation   1476
             /note="G substituted in clone: GSI-180123"
             /replace="T"
variation   1684
             /note="G substituted in clone: CTC-806C5"
             /replace="A"
variation   1699
             /note="T substituted in clone: CTC-806C5"
             /replace="A"
variation   1718
             /note="C substituted in clone: CTC-806C5"
             /replace="G"
variation   1956
             /note="T substituted in clone: CTC-806C5"
             /replace="C"
variation   2838
             /note="T substituted in clone: GSI-180123"
             /replace="C"
variation   3492
             /note="C substituted in clone: CTC-806C5"
             /replace="T"
variation   3795
             /note="T substituted in clone: CTC-806C5"
             /replace="C"
variation   3847
             /note="T substituted in clone: CTC-806C5"
             /replace="C"
variation   3962
             /note="G substituted in clone: CTC-806C5"
             /replace="A"
variation   4097
             /note="A substituted in clone: CTC-806C5"
             /replace="T"
variation   4357

```

```

/note="deleted in clone: GSI-180123"
/replace="G"
4787
/note="A substituted in clone: CTC-806C5"
/replace="C"
5848
/note="C substituted in clone: GSI-180123"
/replace="T"
6653
/note="T substituted in clone: CTC-806C5"
/replace="G"
6828
/note="A substituted in clone: CTC-806C5"
/replace="G"
7941
/note="A substituted in clone: GSI-180123"
/replace="G"
complement(16883..16987)
/note="single stranded/single chemistry region"
complement(63278..63368)
/note="single clone coverage"
complement(63298..63300)
complement(63315..63321)
complement(63323..63325)
complement(63328..63331)
complement(63334..63348)
63365
63368
complement(63369..64555)
/note="single stranded/single chemistry region"
63370..63381
complement(63526..63558)
/note="low quality region"
complement(63527..63536)
65610..65795
/note="single stranded/single chemistry region"
67852
/note="low quality region"
complement(77835..77950)
/note="single stranded/single chemistry region"
complement(79147..79198)
/note="pcr product sequence only , CTC-806C5"
complement(79513..79832)
/note="single stranded/single chemistry region"
complement(79513..79531)
/note="single clone coverage"
complement(79518..79520)
/note="low quality region"
96001
/note="low quality region"
complement(113209..113565)
/note="single stranded/single chemistry region"
ORIGIN
Query Match      6.0%; Score 59; DB 9; Length 122029;
Best Local Similarity 100.0%; Pred No. 2.8e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 CGATCTCTGACCTCGTGATCCGCCGCCCTCAGCTCCCAAGTGTGGATTACAGGC 92
DB 116766 CGATCTCTGACCTCGTGATCCGCCGCCCTCAGCTCCCAAGTGTGGATTACAGGC 116708

```

## RESULT 13

```

AC104070      AC104070      123291 bp      DNA      linear      PRI 29-MAY-2002
LOCUS
DEFINITION Homo sapiens BAC clone RP11-279K24 from 4, complete sequence.
ACCESSION AC104070 AC068461
VERSION AC104070.3 GI:20279508
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

## REFERENCE

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE 1 (bases 1 to 123291)  
 JOURNAL Sulston, J.E. and Waterston, R.  
 MEDLINE Toward a complete human genome sequence  
 PUBMED Genome Res. 8 (11), 1097-1108 (1998)  
 99063792

## REFERENCE

AUTHORS 2 (bases 1 to 123291)  
 TITLE Desai, A., Kozlowski, A. and Boyer, E.  
 JOURNAL The sequence of Homo sapiens BAC clone RP11-279K24  
 MEDLINE Unpublished (2001)  
 PUBMED 3 (bases 1 to 123291)  
 9847074

## REFERENCE

AUTHORS Direct Submission  
 TITLE Waterston, R.H.  
 JOURNAL Submitted (03-DEC-2001) Genome Sequencing Center, Washington  
 MEDLINE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 PUBMED MO 63108, USA  
 40 (bases 1 to 123291)

## REFERENCE

AUTHORS Direct Submission  
 TITLE Waterston, R.H.  
 JOURNAL Submitted (24-APR-2002) Genome Sequencing Center, Washington  
 MEDLINE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 PUBMED MO 63108, USA  
 5 (bases 1 to 123291)

## REFERENCE

AUTHORS Direct Submission  
 TITLE Waterston, R.  
 JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington  
 MEDLINE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 PUBMED On Apr 24, 2002 this sequence version replaced gi:18030153.

## COMMENT

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0279K24  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, S., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-45I20, 2000 bp overlap; the clone sequenced to the right is RP11-173M11. Actual start of this clone is at base position 108871 of RP11-45I20; actual end is at base position 123291 of RP11-279K24.

Unresolved tandem repeats exist between 44681 and 46316.  
 Polymorphisms exist between AC096659, AC0110771 and AC104070. Data from AC110771 was used to finish AC104070.

The sequence of AC068461 has been incorporated into AC104070.

## FEATURES

Source	Location/Qualifiers
1. .123291	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-279K24"
	/clone_lib="RPC1-11"
repeat_region	174..350
	/rpt_family="MER1_type"
repeat_region	1085..1129
	/rpt_family="(CA)n"
repeat_region	2338..3252
	/rpt_family="Alu"
repeat_region	3691..3908
	/rpt_family="L2"
repeat_region	4040..4334
	/rpt_family="Alu"
repeat_region	4345..4522
	/rpt_family="Alu"
repeat_region	4525..5166
	/rpt_family="ERV1"
repeat_region	5269..5343
	/rpt_family="MIR"
repeat_region	5469..5535
	/rpt_family="MIR"
repeat_region	5866..6417
	/rpt_family="ERV1"
repeat_region	7027..7321
	/rpt_family="L2"
repeat_region	7331..7424
	/rpt_family="L2"
repeat_region	7535..7724
	/rpt_family="ERV1"
repeat_region	7739..8079
	/rpt_family="ERV1"
repeat_region	11056..11367
	/rpt_family="Alu"
repeat_region	12482..12789
	/rpt_family="Alu"
repeat_region	12790..12814
	/rpt_family="AT-rich"
repeat_region	13276..13375
	/rpt_family="L2"
repeat_region	13504..13774
	/rpt_family="L1"
repeat_region	13776..14086
	/rpt_family="Alu"
repeat_region	14090..14277
	/rpt_family="L1"
repeat_region	14307..14427
	/rpt_family="L1"
repeat_region	14842..14709
	/rpt_family="U4"
repeat_region	14802..14883
	/rpt_family="CT-rich"
repeat_region	15149..15220
	/rpt_family="ERV1"
repeat_region	15221..15419
	/rpt_family="MaLR"
repeat_region	15420..16124
	/rpt_family="ERV1"
repeat_region	16124..16356
	/rpt_family="MaLR"
repeat_region	16357..16729
	/rpt_family="ERV1"

```

repeat_region 16730..17029
/rpt_family="Alu"
repeat_region 17030..17494
/rpt_family="ERV1"
repeat_region 17517..17607
/rpt_family="L2"
repeat_region 18972..19151
/rpt_family="ERV1"
repeat_region 19204..19453
/rpt_family="ERV1"
repeat_region 19564..19640
/rpt_family="GA-rich"
repeat_region 20193..21044
/rpt_family="ERV1"
repeat_region 21501..21792
/rpt_family="Alu"
repeat_region 23724..23747
/rpt_family="AT-rich"
repeat_region 24122..24419
/rpt_family="Alu"
repeat_region 24525..24920
/rpt_family="MaLR"
repeat_region 25364..25788
/rpt_family="ERV1"
repeat_region 26725..27020
/rpt_family="Alu"
repeat_region 27333..27551
/rpt_family="MIR"
repeat_region 28446..28951
/rpt_family="Alu"
repeat_region 28952..29088
/rpt_family="Alu"
repeat_region 29204..29740
/rpt_family="L1"
repeat_region 29741..30050
/rpt_family="Alu"
repeat_region 30051..30244
/rpt_family="L1"
repeat_region 30245..30462
/rpt_family="Alu"
repeat_region 30463..30556
/rpt_family="L1"
repeat_region 30733..30767
/rpt_family="U2"
repeat_region 30768..30845
/rpt_family="L1"

Query Match 6.0%; Score 59; DB 9; Length 123291;
Best Local Similarity 100.0%; Pred. No. 2.8e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCCTGACCTCGATCGCGCCCTCAGCTTCCCAAGTGTGGGATTACAGGC 92
|||||
Db 43370 CGATCTCCTGACCTCGATCGCGCCCTCAGCTTCCCAAGTGTGGGATTACAGGC 43428

RESULT 14
BX571818/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-39J22, WORKING DRAFT SEQUENCE,
5 unordered pieces.
ACCESSION BX571818
VERSION BX571818.2 GI:33386608
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148963)
McLay, K.
Direct Submission
TITLE
JOURNAL Submitted (30-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

```

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Jul 31, 2003 this sequence version replaced gi:33086351.

## COMMENT

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA38J22  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 147613 bases at least Q40  
 Consensus quality: 147851 bases at least Q30  
 Consensus quality: 148031 bases at least Q20  
 Insert size: 148563; sum-of-contigs  
 Insert size: 148899; 19.7% error; agarose-fp  
 Quality coverage: 24.62x in Q20 bases; sum-of-contigs Quality  
 coverage: 25.09x in Q20 bases; agarose-fp

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2276: contig of 2276 bp in length  
 \* 2277 2376: gap of 100 bp  
 \* 2377 35686: contig of 33310 bp in length  
 \* 35687 35786: gap of 100 bp  
 \* 35787 51485: contig of 15699 bp in length  
 \* 51486 51585: gap of 100 bp  
 \* 51586 132546: contig of 80961 bp in length  
 \* 132547 132646: gap of 100 bp  
 \* 132647 148963: contig of 16317 bp in length.

## FEATURES

source

1..148963  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-38J22"  
 /clone\_lib="RPC1-11.1"  
 1..2276  
 /note="assembly fragment:00367"

misc\_feature

fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left  
 2377..35686  
 /note="assembly fragment:04519"

misc\_feature

fragment\_chain:1  
 35787..51485  
 /note="assembly fragment:03291"

misc\_feature

fragment\_chain:1  
 51586..132546  
 /note="assembly fragment:03347"

## ORIGIN

Query Match 6.0%; Score 59; DB 2; Length 148963;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-21;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GTGATCCCGCCGCTCAGCTTCCCAAGTGTGGGATTACAGGCACCGCC 107  
 |||||  
 Db 15354 GTGATCCCGCCGCTCAGCTTCCCAAGTGTGGGATTACAGGCACCGCC 15296



RESULT 15  
AC012440  
LOCUS  
DEFINITION Homo sapiens clone RP11-11H11, WORKING DRAFT SEQUENCE, 21 unordered  
pieces.  
AC012440  
VERSION AC012440.2 GI:7137115  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 159619)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone RP11-11H11  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 159619)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckghalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeAvaliano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye-S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
DIRECT SUBMISSION  
SUBMITTED (27-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
ON MAR 1, 2000 this sequence version replaced gi:6136376.  
ALL REPEATS were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center Project name: L3166  
Center Clone name: 11 H.11  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 149892 bases at least Q40  
Consensus quality: 154478 bases at least Q30  
Consensus quality: 156029 bases at least Q20  
Insert size: 165000; agarose-fp  
Insert size: 157619; sum-of-contigs  
Quality coverage: 4.7 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1125: contig of 1125 bp in length  
\* 1126 1225: gap of 100 bp  
\* 1226 2918: contig of 1693 bp in length  
\* 2919 3019: gap of 100 bp  
\* 3019 4446: contig of 1428 bp in length

4447 4546: gap of 100 bp  
4547 5800: contig of 1254 bp in length  
5801 5901: gap of 100 bp  
5901 6771: contig of 871 bp in length  
6772 6871: gap of 100 bp  
6872 8592: contig of 1721 bp in length  
8593 8692: gap of 100 bp  
8693 11437: contig of 2745 bp in length  
11438 11537: gap of 100 bp  
11538 14484: contig of 2947 bp in length  
14485 14584: gap of 100 bp  
14585 17422: contig of 2838 bp in length  
17423 17523: gap of 100 bp  
17524 21714: contig of 4192 bp in length  
21715 21814: gap of 100 bp  
21815 27645: contig of 5832 bp in length  
27646 27747: gap of 100 bp  
27748 33347: contig of 5601 bp in length  
33348 33447: gap of 100 bp  
33448 41489: contig of 8042 bp in length  
41490 41590: gap of 100 bp  
41591 49388: contig of 8398 bp in length  
49389 50087: gap of 100 bp  
50088 57307: contig of 7219 bp in length  
57308 57406: gap of 100 bp  
57407 63803: contig of 6397 bp in length  
63804 63903: gap of 100 bp  
63904 74106: contig of 10203 bp in length  
74107 74206: gap of 100 bp  
74207 85991: contig of 11785 bp in length  
85992 86092: gap of 100 bp  
86093 104013: contig of 17921 bp in length  
104014 104113: gap of 100 bp  
104114 124486: contig of 20374 bp in length  
124487 124586: gap of 100 bp  
124587 159619: contig of 35033 bp in length.

FEATURES  
source  
1..159619  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-11H11"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..1125  
/note="assembly\_fragment"  
1226..2918  
/note="assembly\_fragment"  
3019..4446  
/note="assembly\_fragment"  
4547..5800  
/note="assembly\_fragment"  
5901..6771  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
6872..8592  
/note="assembly\_fragment"  
8693..11437  
/note="assembly\_fragment"  
11538..14484  
/note="assembly\_fragment"  
14585..17422  
/note="assembly\_fragment"  
17523..21714  
/note="assembly\_fragment"  
21815..27646  
/note="assembly\_fragment"  
27747..33347  
/note="assembly\_fragment"  
33448..41489  
/note="assembly\_fragment"  
41590..49987  
/note="assembly\_fragment"

```

misc_feature      50088..57306
                  /note="assembly_fragment"
misc_feature      57407..63803
                  /note="assembly_fragment"
misc_feature      63904..74106
                  /note="assembly_fragment"
misc_feature      74207..85991
                  /note="assembly_fragment"
misc_feature      86092..104012
                  /note="assembly_fragment"
misc_feature      104113..124486
                  /note="assembly_fragment"
                  /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:left"
misc_feature      124587..159619
                  /note="assembly_fragment"

ORIGIN
Query Match      6.0%; Score 59; DB 2; Length 159619;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      49 GTGATCGCGCGCTCAGCTTCCCAAGTGGTGGGATTACAGGCATGAGCCACCGCGCC 107
Db      75510 GTGATCGCGCGCTCAGCTTCCCAAGTGGTGGGATTACAGGCATGAGCCACCGCGCC 75568

```

Search completed: July 10, 2004, 14:37:04  
Job time : 4191 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 13:09:30 ; Search time 99 Seconds  
(without alignments)  
5549.509 Million cell updates/sec

Title: US-09-972-032-1  
Perfect score: 990  
Sequence: 1 ggaatgttctcgaggccaa.....aaatgagcgcgcaagt 990

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	4.8	118067	4	US-09-497-855A-32
2	45	4.5	505	4	US-09-621-976-15373
3	45	4.5	3267	2	US-08-257-963B-12
4	45	4.5	3267	4	US-08-367-841A-12
5	45	4.5	3267	5	PCT-US95-07201-12
6	45	4.5	5262	4	US-08-520-373D-5
7	45	4.5	22481	4	US-08-367-841A-43
8	45	4.5	22481	5	PCT-US95-07201-43
9	45	4.5	22484	4	US-09-875-223-2
10	45	4.5	22484	4	US-09-875-114-2
11	45	4.5	75395	4	US-09-984-890-3
12	44	4.4	164	4	US-09-621-976-11863
13	44	4.4	281	4	US-09-621-976-1944
14	44	4.4	294	4	US-09-621-976-1939
15	44	4.4	294	4	US-09-621-976-1940
16	44	4.4	298	4	US-09-621-976-11526
17	44	4.4	298	4	US-09-621-976-13327
18	44	4.4	302	4	US-09-621-976-1937
19	44	4.4	302	4	US-09-621-976-1942
20	44	4.4	310	4	US-09-621-976-1943
21	44	4.4	312	4	US-09-621-976-12169
22	44	4.4	314	4	US-09-621-976-11935
23	44	4.4	314	4	US-09-621-976-11935
24	44	4.4	507	4	US-09-621-976-13632
25	44	4.4	99500	4	US-09-788-096-10
26	44	4.4	392000	4	US-10-027-983-11
27	43	4.3	4858	4	US-09-595-684B-28

C 28	43	4.3	14364	4	US-10-067-443-20	Sequence 20, Appl
C 29	43	4.3	786431	4	US-09-751-389-3	Sequence 3, Appl
C 30	42	4.2	3350	3	US-09-110-116-2	Sequence 2, Appl
C 31	42	4.2	21784	4	US-09-820-002-3	Sequence 3, Appl
C 32	41	4.1	75395	4	US-09-984-890-3	Sequence 3, Appl
C 33	40	4.0	421	4	US-09-621-976-15130	Sequence 15130, A
C 34	40	4.0	438	4	US-09-621-976-15688	Sequence 15688, A
C 35	40	4.0	7210	2	US-08-257-963B-10	Sequence 10, Appl
C 36	40	4.0	7210	4	US-08-367-841A-10	Sequence 10, Appl
C 37	40	4.0	7210	5	PCT-US95-07201-10	Sequence 10, Appl
C 38	40	4.0	11811	3	US-09-078-294-7	Sequence 7, Appl
C 39	40	4.0	14581	4	US-08-520-373D-4	Sequence 4, Appl
C 40	40	4.0	20303	1	US-08-370-975B-6	Sequence 6, Appl
C 41	40	4.0	26764	1	US-08-370-975B-1	Sequence 1, Appl
C 42	40	4.0	70000	4	US-09-851-896-3	Sequence 3, Appl
C 43	40	4.0	112132	4	US-09-741-150-3	Sequence 3, Appl
C 44	40	4.0	112132	4	US-10-160-187-3	Sequence 3, Appl
C 45	40	4.0	152331	3	US-09-128-155-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-497-855A-32

; Sequence 32, Application US/09497855A

; Patent No. 6605432

; GENERAL INFORMATION:

; APPLICANT: Huang, Tim

; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION

; FILE REFERENCE: UM01523

; CURRENT APPLICATION NUMBER: US/09/497,855A

; CURRENT FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: 60/120,592

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: 60/118,760

; PRIOR FILING DATE: 1999-02-05

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 32

; LENGTH: 118067

; TYPE: DNA

; ORGANISM: Homo sapiens;

US-09-497-855A-32

Query Match 4.8%; Score 48; DB 4; Length 118067;

Best Local Similarity 100.0%; Pred. No. 1.2e-11;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GTGATCCGCCGCCCTCAGCTCCCAAGTGTGGGATTACAGGCATGA 96

DB 33745 GTGATCCGCCGCCCTCAGCTCCCAAGTGTGGGATTACAGGCATGA 33792

RESULT 2

US-09-621-976-15373

; Sequence 15373, Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 15373

; LENGTH: 506

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-621-976-15373

Query Match 4.5%; Score 46; DB 4; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 114  
 |||||  
 Db 76 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 121  
 |||||

## RESULT 3

US-08-257-963B-12  
 ; Sequence 12, Application US/08257963B  
 ; Patent No. 5840686  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chader, Gerald J.; Becerra, S.  
 ; APPLICANT: Patricia; Schwartz, Joan P.;  
 ; APPLICANT: Taniwaki, Takayuki  
 ; TITLE OF INVENTION: PIGMENT EPITHELIUM  
 ; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL  
 ; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING  
 ; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/257,963B  
 ; FILING DATE: 24-SEP-1992  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/952,796  
 ; FILING DATE: 24-SEP-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: DOROTHY R. AUTH  
 ; REGISTRATION NUMBER: 36434  
 ; REFERENCE/DOCKET NUMBER: 20264126US1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 751-6849  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3267 Base Pairs  
 ; TYPE: Nucleic Acid  
 ; STRANDEDNESS: Double  
 ; TOPOLOGY: Unknown  
 ; MOLECULE TYPE: Genomic DNA  
 ; FEATURE:  
 ; NAME/KEY: JT109  
 ; LOCATION:  
 ; IDENTIFICATION METHOD:  
 ; OTHER INFORMATION: 3.3 kb PCR product  
 ; OTHER INFORMATION: using primers, SEQ ID No. 5840686 15 and 16  
 US-08-257-963B-12

Query Match 4.5%; Score 45; DB 2; Length 3267;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113  
 |||||  
 Db 625 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669  
 |||||

## RESULT 4

US-08-367-841A-12  
 ; Sequence 12, Application US/08367841A  
 ; Patent No. 6319687  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chader, Gerald J.; Rodriguez,  
 ; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;  
 ; APPLICANT: Tombran-Tink, Joyce  
 ; TITLE OF INVENTION: PIGMENT EPITHELIUM  
 ; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC  
 ; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/367,841A  
 ; FILING DATE: 30-DEC-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/257,963  
 ; FILING DATE: 07-JUN-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/952,796  
 ; FILING DATE: 24-SEP-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: DOROTHY R. AUTH  
 ; REGISTRATION NUMBER: 36434  
 ; REFERENCE/DOCKET NUMBER: 20264126US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 751-6849  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3267 Base Pairs  
 ; TYPE: Nucleic Acid  
 ; STRANDEDNESS: Double  
 ; TOPOLOGY: Unknown  
 ; MOLECULE TYPE: Genomic DNA  
 ; FEATURE:  
 ; NAME/KEY: JT109  
 ; LOCATION:  
 ; IDENTIFICATION METHOD:  
 ; OTHER INFORMATION: 3.3 kb PCR product  
 ; OTHER INFORMATION: using primers, SEQ ID No. 6319687 15 and 16  
 US-08-367-841A-12

Query Match 4.5%; Score 45; DB 4; Length 3267;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113  
 |||||  
 Db 625 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669  
 |||||

## RESULT 5

PCT-US95-07201-12  
 ; Sequence 12, Application PC/TUS9507201  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chader, Gerald J.; Becerra, Sofia  
 ; APPLICANT: Patricia; Schwartz, Joan P.;  
 ; APPLICANT: Taniwaki, Takayuki  
 ; TITLE OF INVENTION: PIGMENT EPITHELIUM

TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC  
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10154

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07201

FILING DATE: 06-JUN-1995

## CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/367,841

FILING DATE: 30-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/257,963

FILING DATE: 07-JUN-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/952,796

FILING DATE: 24-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: DOROTHY R. AUTH

REGISTRATION NUMBER: 36434

REFERENCE/POCKET NUMBER: 20264126PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 3267 Base Pairs

TYPE: Nucleic Acid

STRANDEDNESS: Double

TOPOLOGY: Unknown

MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: JT109

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION: 3.3 kb PCR product

OTHER INFORMATION: using primers, SEQ ID No: 15 and 16

PCT-US95-07201-12

Query Match 4.5%; Score 45; DB 5; Length 3267;

Best Local Similarity 100.0%; Pred. No. 2.8e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGCATGAGCCACCGCGCGGCC 113

DB 625 TCCCAAAGTGTGGATTACAGGCATGAGCCACCGCGCGGCC 669

## RESULT 6

US-08-520-373D-5

Sequence 5, Application US/08520373D

Patent No. 6451763

GENERAL INFORMATION:

APPLICANT: Tombran-Tink, Joyce

APPLICANT: Steele, Fintan R

APPLICANT: Chader, Gerald J

APPLICANT: Becerra, Sofia P

APPLICANT: Johnson, Lincoln V

APPLICANT: Rodriguez, Ignacio R

TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR

FILE REFERENCE: 2026-4203US1

CURRENT APPLICATION NUMBER: US/08/520,373D

CURRENT FILING DATE: 1995-08-29  
PRIOR APPLICATION NUMBER: 08/377,710  
PRIOR FILING DATE: 1995-01-25  
PRIOR APPLICATION NUMBER: 08/279,979  
PRIOR FILING DATE: 1994-07-25  
PRIOR APPLICATION NUMBER: 07/894,215  
PRIOR FILING DATE: 1992-06-04  
PRIOR APPLICATION NUMBER: 07/952,796  
PRIOR FILING DATE: 1992-09-24  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5

LENGTH: 5262

TYPE: DNA

ORGANISM: HUMAN

FEATURE:

OTHER INFORMATION: EXON 35-161; EXON 1142-1297; EXON 1984-2187;

OTHER INFORMATION: EXON 5170-5255; INTRON 162-1141; INTRON

OTHER INFORMATION: 1298-1983; INTRON 2188-5169; CDS 35-161; CDS

OTHER INFORMATION: 1142-1297; CDS 1984-2187; CDS 5170-5255

NAME/KEY: exon

LOCATION: (35)..(160)

NAME/KEY: exon

LOCATION: (1142)..(1297)

NAME/KEY: exon

LOCATION: (1984)..(2187)

NAME/KEY: exon

LOCATION: (5170)..(5256)

NAME/KEY: intron

LOCATION: (162)..(1141)

NAME/KEY: intron

LOCATION: (1298)..(1983)

NAME/KEY: intron

LOCATION: (2188)..(5169)

OTHER INFORMATION: n = a or g or t or c, any base

US-08-520-373D-5

Query Match 4.5%; Score 45; DB 4; Length 5262;

Best Local Similarity 100.0%; Pred. No. 2.7e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGCATGAGCCACCGCGCGGCC 113

DB 2619 TCCCAAAGTGTGGATTACAGGCATGAGCCACCGCGCGGCC 2663

## RESULT 7

US-08-367-841A-43

Sequence 43, Application US/08367841A

Patent No. 6319587

GENERAL INFORMATION:

APPLICANT: Chader, Gerald J.; Rodriguez,

APPLICANT: Ignacio R.; Mazuruk, Krzysztof;

APPLICANT: Tombran-Tink, Joyce

TITLE OF INVENTION: PIGMENT EPITHELIUM

TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC

TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan & Finnegan

STREET: 345 Park Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/367,841A

FILING DATE: 30-DEC-1994

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,963  
FILING DATE: 07-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/952,796  
FILING DATE: 24-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DOROTHY R. AUTH  
REGISTRATION NUMBER: 36434  
REFERENCE/DOCKET NUMBER: 20264126US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22481 Base Pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Double  
TOPOLOGY: Unknown  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Pl-147  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: full length genomic  
OTHER INFORMATION: sequence for PEDF plus flanking sequences.  
US-08-367-841A-43

Query Match 4.5%; Score 45; DB 4; Length 22481;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113  
Db 17186 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 8  
PCT-US95-07201-43  
Sequence 43, Application PC/TUS9507201  
GENERAL INFORMATION:  
APPLICANT: Chader, Gerald J.; Becerra, Sofia  
APPLICANT: Patricia; Schwartz, Joan P.;  
APPLICANT: Taniwaki, Takayuki  
TITLE OF INVENTION: PIGMENT EPITHELIUM  
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC  
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07201  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/367,841  
FILING DATE: 30-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,963  
FILING DATE: 07-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/952,796

FILING DATE: 24-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DOROTHY R. AUTH  
REGISTRATION NUMBER: 36434  
REFERENCE/DOCKET NUMBER: 20264126PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22481 Base Pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Double  
TOPOLOGY: Unknown  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Pl-147  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: full length genomic  
OTHER INFORMATION: sequence for PEDF plus flanking sequences.  
PCT-US95-07201-43

Query Match 4.5%; Score 45; DB 5; Length 22481;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113  
Db 17186 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 9  
US-09-875-223-2  
Sequence 2, Application US/09875223  
Patent No. 6391850  
GENERAL INFORMATION:  
APPLICANT: No. 6391850thwestern University  
APPLICANT: No. 63918501 Bouck  
APPLICANT: David Dawson  
APPLICANT: Paul Gillis  
TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis  
FILE REFERENCE: 0290-23U3  
CURRENT APPLICATION NUMBER: US/09/875,223  
CURRENT FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 09/122,079  
PRIOR FILING DATE: 1998-07-23  
PRIOR APPLICATION NUMBER: PCT/US98/15228  
PRIOR FILING DATE: 1998-07-23  
PRIOR APPLICATION NUMBER: US 08/899,304  
PRIOR FILING DATE: 1997-07-23  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 22484  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: 1...22484  
OTHER INFORMATION: "n" means either a, c, t, or g  
US-09-875-223-2

Query Match 4.5%; Score 45; DB 4; Length 22484;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113  
Db 17186 TCCCAAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 10

```
US-09-875-114-2
; Sequence 2, Application US/09875114
; Patent No. 6670333
; GENERAL INFORMATION:
; APPLICANT: No. 6670333thwestern University
; APPLICANT: No. 66703331 Bouck
; APPLICANT: David Dawson
; APPLICANT: Paul Willis
; TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
; FILE REFERENCE: 0290-2302
; CURRENT APPLICATION NUMBER: US/09/875,114
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 09/122,079
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: PCT/US98/15228
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: US 08/899,304
; PRIOR FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 22484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: 1...22484
; OTHER INFORMATION: "n" means either a, c, t, or g
US-09-875-114-2

Query Match      4.5%; Score 45; DB 4; Length 22484;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 113
Db 17186 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 11
US-09-984-890-3/c
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001306
; CURRENT APPLICATION NUMBER: US/09/984,890
; CURRENT FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 75395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(75395)
; OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

Query Match      4.5%; Score 45; DB 4; Length 75395;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 113
Db 12402 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 12358

RESULT 12
```

```
US-09-621-976-11863/c
; Sequence 11863, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 11863
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-621-976-11863

Query Match      4.4%; Score 44; DB 4; Length 164;
Best Local Similarity 100.0%; Pred. No. 9.2e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 112
Db 59 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 16

RESULT 13
US-09-621-976-1944/c
; Sequence 1944, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1944
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..195
; NAME/KEY: sig_peptide
; LOCATION: 34..189
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 3.7999995231628
; OTHER INFORMATION: seq LGAVAGACSPSCS/GG
US-09-621-976-1944

Query Match      4.4%; Score 44; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 112
Db 59 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 16

RESULT 14
US-09-621-976-1939/c
; Sequence 1939, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
```

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 1939

; LENGTH: 294

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34..195

; NAME/KEY: sig\_peptide

; LOCATION: 34..189

; OTHER INFORMATION: Von Heijne matrix

; OTHER INFORMATION: score 3.79999995231628

; OTHER INFORMATION: seq LGAVAGACSPSCS/GG

US-09-621-976-1939

Query Match

Best Local Similarity 4.4%; Score 44; DB 4; Length 294;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 112

Db 59 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 16

RESULT 15

US-09-621-976-1940/c

; Sequence 1940. Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 1940

; LENGTH: 294

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34..195

; NAME/KEY: sig\_peptide

; LOCATION: 34..189

; OTHER INFORMATION: Von Heijne matrix

; OTHER INFORMATION: score 3.79999995231628

; OTHER INFORMATION: seq LGAVAGACSPSCS/GG

US-09-621-976-1940

Query Match

Best Local Similarity 4.4%; Score 44; DB 4; Length 294;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 112

Db 59 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 16

Search completed: July 10, 2004, 15:30:57

Job time : 103 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 13, 2004, 08:38:15 ; Search time 375 Seconds  
(without alignments)  
894.954 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 79  
Sequence: 1 MCGRRVRSAGCGFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 6744130

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2 1/USPTO.spool/US09972032/runat 06072004 121533 8862/app\_query.fasta\_1.263  
-DB=N Geneseq 29Jan04 -OFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972032 @CGN 1.1 470 @runat 06072004 121533 8862 -NCFU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002s.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	79	100.0	990	ABL60606	ABL60606 Human ERC
C 2	26	32.9	597	AAS69040	Aas69040 DNA encod
C 3	9	11.4	1402	AAV36085	AAV36085 DNA encod
C 4	8	10.1	379	AAF64670	AAF64670 Novel hum
C 5	8	10.1	483	ABZ37949	Abz37949 N. gonorr
C 6	8	10.1	600	AA65458	Aa65458 Porcine B
C 7	8	10.1	720	ACA23355	Aca23355 Prokaryot
C 8	8	10.1	1015	AAF12681	Aaf12681 Aspergill

C 9	8	10.1	1035	7	ACA25574	Aca25574 Prokaryot
C 10	8	10.1	1037	4	AAH48630	Aah48630 Human MLP
C 11	8	10.1	1230	6	ABQ73845	Abq73845 Rhizobium
C 12	8	10.1	1239	7	ACA25997	Aca25997 Prokaryot
C 13	8	10.1	1266	6	ABQ73844	Abq73844 Yersinia
C 14	8	10.1	1349	2	AAQ06668	Aaq06668 Mycomacte
C 15	8	10.1	1394	6	ABQ79295	Abq79295 Mycobacte
C 16	8	10.1	1396	6	ABK52535	Abk52535 AIDS/hepa
C 17	8	10.1	1437	4	AAQ10372	Aaq10372 Tuberculi
C 18	8	10.1	1437	4	AAS53940	Aas53940 Klebstell
C 19	8	10.1	1536	7	ACA35721	Aca35721 Prokaryot
C 20	8	10.1	1680	4	ABL26019	AbL26019 Drosophil
C 21	8	10.1	1845	7	ACC43487	Acc43487 Nucleotid
C 22	8	10.1	1914	6	ABK51971	Abk51971 Corn cDNA
C 23	8	10.1	1914	7	ABX93207	Abx93207 cDNA enco
C 24	8	10.1	2046	7	ACA28541	Aca28541 Prokaryot
C 25	8	10.1	2169	7	ADA53246	Ada53246 Human cod
C 26	8	10.1	2265	4	ABL09455	AbL09455 Drosophil
C 27	8	10.1	2391	4	AAK77091	Aak77091 Human imm
C 28	8	10.1	2728	6	ABQ99435	Abq99435 Human cod
C 29	8	10.1	2883	4	ABL25994	AbL25994 Drosophil
C 30	8	10.1	2929	6	ABK81822	Abk81822 DNA repre
C 31	8	10.1	3680	4	ABL26018	AbL26018 Drosophil
C 32	8	10.1	3752	3	AAA65427	Aaa65427 Porcine B
C 33	8	10.1	5877	6	ABS78681	Abs78681 Kitasatos
C 34	8	10.1	9857	4	ABL09454	AbL09454 Drosophil
C 35	8	10.1	21034	2	AAV62154	Aav62154 HSV-2 str
C 36	8	10.1	22428	4	AA541759	Aa541759 Genomic s
C 37	8	10.1	22428	7	ABZ67767	Abz67767 Human sec
C 38	8	10.1	22428	7	ADA98730	Ada98730 Human sec
C 39	8	10.1	26338	2	AAV62134	Aav62134 HSV-2 str
C 40	8	10.1	36955	6	ABV73608	Abv73608 S. albulu
C 41	8	10.1	49999	2	AAZ23302	Aaz23302 Human LOB
C 42	8	10.1	92934	3	AA81473	Aa81473 N. mening
C 43	8	10.1	110000	3	AA81489	Continuation (8 of
C 44	8	10.1	110000	3	AA81489	Continuation (8 of
C 45	8	10.1	110000	4	AA199682	Continuation (10 of

ALIGNMENTS

RESULT 1  
ABL60606  
ID ABL60606 standard; cDNA; 990 BP.

XX ABL60606;  
AC ABL60606;  
XX 27-AUG-2002 (first entry)  
DT 27-AUG-2002 (first entry)  
XX Human ERCoA3 protein encoding cDNA.  
DE Human ERCoA3 protein encoding cDNA.  
XX Btrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;  
KW osteoporosis; cytostatic; osteopathic; human; gene; ss.  
XX Homo sapiens.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 203..442  
FT /\*tag= a  
FT /\*product= "ERCoA3 protein"  
XX WO200228352-A2.

XX 11-APR-2002.  
XX 05-OCT-2001; 2001WO-US031271.  
XX 05-OCT-2000; 2000US-0238190P.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX Montano M, Sutton A;  
XX



PN US5776692-A.  
 XX 07-JUL-1998.  
 PD 23-MAY-1995; 95US-00447965.  
 XX 23-MAY-1995; 95US-00447965.  
 PF (BAVU ) BAYLOR COLLEGE MEDICINE.  
 XX Graham DY, El-Zaatari FAK, Naser S;  
 PI WPI; 1998-398024/34.  
 XX P-PSDB; AAW60723.  
 DR Recombinant clone encoding new Mycobacterium paratuberculosis protein -  
 PT containing nucleic acid useful as genus-specific hybridisation probe for  
 PT detecting mycobacteria.  
 XX Claim 2; Fig 5; 12pp; English.  
 PS The present sequence represents the BamHI-DNA insert of the recombinant  
 CC clone pMpb #48 and encodes a 36K antigen of Mycobacterium  
 CC paratuberculosis. The DNA sequence and its fragments are useful as  
 CC hybridisation probes and amplification primers for detecting nucleic acid  
 CC specific for the genus Mycobacterium, particularly for diagnosis of  
 CC infection, especially by M. paratuberculosis which has been implicated in  
 CC Crohn's disease and sarcoidosis. The 36K protein is useful in serological  
 CC diagnosis and for vaccine development. (Updated on 17-OCT-2003 to  
 CC standardise OS field)  
 XX SQ Sequence 1402 BP; 171 A; 490 C; 552 G; 185 T; 0 U; 4 Other;  
 Alignment Scores:  
 Pred. No.: 35-7 Length: 1402  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.39% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-972-032-2 (1-79) x AAV36085 (1-1402)  
 Qy 47 AlasArgTrpProArgSerAlaSer 55  
 Db 1260 GCGTCCCGTTGGCCGAGATCGGCAGT 1234  
 RESULT 4  
 AAF64670/C  
 ID AAF64670 standard; cDNA; 379 BP.  
 XX AC AAF64670;  
 XX 09-APR-2001 (first entry)  
 XX DE Novel human polynucleotide, SEQ ID NO: 426.  
 XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
 KW breast cancer; lung cancer; cancer detection; ss.  
 XX Homo sapiens.  
 XX OS WO2001025568-A2.  
 XX PN WO2001025568-A2.  
 XX PD 11-JAN-2001.  
 XX 30-JUN-2000; 2000WO-US018374.  
 XX PF 02-JUL-1999; 99US-0142310P.  
 PR 02-JUL-1999; 99US-0142311P.  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;  
 PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;  
 PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;  
 XX WPI; 2001-091805/10.  
 DR Library of polynucleotides for diagnosing a cancerous state of a  
 XX mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 3351 human polynucleotide sequences.  
 XX Claim 9; Page 605-606; 1046pp; English.  
 PS The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia  
 XX SQ Sequence 379 BP; 79 A; 108 C; 110 G; 82 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 95-4 Length: 379  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 5 Gaps: 0  
 US-09-972-032-2 (1-79) x AAF64670 (1-379)  
 Qy 35 GlyProGluGlyGlnAlaSerPro 42  
 Db 30 GGGCCTGAGGACAGGCCAGTCCT 7  
 RESULT 5  
 ABZ37949/C  
 ID ABZ37949 standard; DNA; 483 BP.  
 XX AC ABZ37949;  
 XX 07-MAR-2003 (first entry)  
 XX DE N. gonorrhoeae nucleotide sequence SEQ ID 487.  
 XX Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 KW Neisseria gonorrhoeae.  
 XX OS WO200279243-A2.  
 XX PN WO200279243-A2.  
 XX PD 10-OCT-2002.  
 XX 12-FEB-2002; 2002WO-IB002069.  
 XX 12-FEB-2001; 2001GB-00003424.  
 XX (CHIR-) CHIRON SPA.  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 PI WPI; 2003-058415/05.  
 DR

DR P-PSDB; ABP76979.  
 XX  
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection.  
 XX  
 XX Disclosure; Page 224; 815pp; English.  
 XX  
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records AB237706-AB242016 represent nucleic acid  
 CC molecules of the invention  
 XX  
 SQ Sequence 483 BP; 110 A; 172 C; 124 G; 77 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 119 Length: 483  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 7 Gaps: 0  
 US-09-972-032-2 (1-79) x AB237949 (1-483)  
 QY 7 ArgValSerAlaGlyCysGlyPhe 14  
 |||||  
 DB 442 CGTGATCAGCAGGATGCTTTT 419  
 RESULT 6  
 AAA65458/c  
 ID AAA65458 standard; DNA; 600 BP.  
 XX  
 AC AAA65458;  
 XX  
 DT 10-NOV-2000 (first entry)  
 XX  
 DE Porcine BAC-PIGF2-1 contig 49.  
 XX  
 KW Porcine; pig; wild boar; quantitative trait locus; QTL; chromosome 2;  
 KW mapping; 2p1.7; select breeding; genotype, phenotype; muscle mass;  
 KW fat deposition; IGF2; insulin-like growth factor 2; ds.  
 XX  
 OS Sus scrofa.  
 XX  
 PN WO200036143-A2.  
 XX  
 PD 22-JUN-2000.  
 XX  
 PF 16-DEC-1999; 98WO-EP010209.  
 XX  
 PR 16-DEC-1998; 98EP-00204291.  
 XX  
 PA (UJLI-) UNIV LIEGE.  
 PA (MELI-) MELICH HB.  
 PA (SEGH-) SEGHERSGENTEC NV.  
 XX  
 PI Andersson L, Georges M, Spincemaille G;  
 XX  
 DR WPI; 2000-431612/37.  
 XX  
 PT Selecting a domestic animal for having desired genotypic properties  
 PT comprises testing the animal for the presence of a parentally imprinted  
 PT quantitative trait locus which is related to muscle mass and/or fat  
 PT deposition.  
 XX  
 PS Example 3; Fig 6; 107pp; English.  
 XX  
 CC The present invention describes a method (M1) for selecting a domestic  
 CC animal for having desired genotypic properties. The method comprises

CC testing the animal for the presence of a parentally imprinted  
 CC quantitative trait locus (QTL). The pig QTL is located at chromosome 2,  
 CC mapping at around position 2p1.7. Also described are: (1) an isolated  
 CC and/or recombinant nucleic acid (N1) comprising a parentally imprinted  
 CC QTL or its functional fragment; (2) an isolated and/or recombinant  
 CC nucleic acid (N2) comprising a synthetic parentally imprinted QTL derived  
 CC from at least one chromosome or its functional fragment; (3) an animal  
 CC such as pig selected for having desired genotypic or potential phenotypic  
 CC properties; (4) a transgenic animal comprising N1 or N2; and (5) sperm or  
 CC an embryo derived from the animal of (3) or (4). N1 or its fragment is  
 CC useful for selecting an animal destined for slaughter or a breeding  
 CC animal having desired genotypic or potential phenotypic properties. The  
 CC properties are related to muscle mass and/or fat deposition. The sperm or  
 CC an embryo are useful in breeding animals destined for slaughter. AAA65418  
 CC to AAA65524 represent contigs 1 to 10 and 19 to 115 which were isolated  
 CC from porcine BAC-PIGF2-1 which contains the INS and IGF2 (insulin-like  
 CC growth factor) genes. These sequences were used in an example from the  
 CC present invention for generating a reference sequence of IGF2 and  
 CC flanking loci in the pig  
 XX  
 SQ Sequence 600 BP; 102 A; 162 C; 228 G; 108 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 146 Length: 600  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 3 Gaps: 0  
 US-09-972-032-2 (1-79) x AAA65458 (1-600)  
 QY 38 GlyGlnAlaSerProThrProAsp 45  
 |||||  
 DB 365 GGCCAAGCGTCACCGACCCCTGAC 342  
 RESULT 7  
 ACA23355/c  
 ID ACA23355 standard; DNA; 720 BP.  
 XX  
 AC ACA23355;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #5012.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS *Borrelia cepacia*.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU19485.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 11225; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 5213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: the sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 720 BP; 127 A; 259 C; 233 G; 101 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 173 Length: 720  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 7 Gaps: 0

US-09-972-032-2 (1-79) x ACA23355 (1-720)

Qy 47 AlaSerArgTrpProArgSerAla 54  
CGGAGCGCGTGGCGCGTTCGGCC 84

Db 107 CGGAGCGCGTGGCGCGTTCGGCC 84

RESULT 8  
AAFI2681  
ID AAF12681 standard; cDNA; 1015 BP.  
AC AAF12681;  
XX  
XX 13-MAR-2001 (first entry)  
DE  
DE Aspergillus oryzae EST SEQ ID NO:5204.  
KW Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX Aspergillus oryzae.  
OS  
XX WO2000056762-A2.  
PN  
XX 28-SEP-2000.  
PD

XX 22-MAR-2000; 2000WO-US007781.  
XX  
XX 22-MAR-1999; 99US-00273623.  
XX  
XX (NOVO) NOVO NORDISK BIOTECH INC.  
PA (NOVO) NOVO NORDISK AS.  
XX  
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
PI WPI; 2000-594572/56.  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags.  
XX  
XX Claim 88; Page 2176-2177; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring the  
CC global expression of genes from FF cells allows the production potential  
CC of the microorganisms to be improved. New genes may be discovered,  
CC possible functions of unknown open reading frames can be identified and  
CC gene copy number variation and stability can be monitored. The expression  
CC of genes can be used to study how FF cells adapt to changes in culture  
CC conditions, environmental stress, spore morphogenesis, recombination,  
CC metabolic or catabolic pathway engineering. Using ESTs provides several  
CC advantages over genomic or random cDNA clones including elimination of  
CC redundancy as one spot on an array equals one gene or open reading frame,  
CC and organisation of the microarrays based on function of the gene  
CC products to facilitate analysis of the results. AAF07478 to AAF11247  
CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents  
CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from  
CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from  
CC *Trichoderma reesei*, which are all specifically claimed in the present  
CC invention  
XX  
SQ Sequence 1015 BP; 222 A; 313 C; 249 G; 230 T; 0 U; 1 Other;

Alignment Scores:  
Pred. No.: 238 Length: 1015  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.13% Indels: 0  
DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x AAF12681 (1-1015)

Qy 49 ArgTrpProArgSerAlaSerArg 56  
CGTGGCGCGGAGCGCGTCCCGG 625

Db 602 CGTGGCGCGGAGCGCGTCCCGG 625

RESULT 9  
ACA25574/c  
ID ACA25574 standard; DNA; 1035 BP.  
XX  
XX ACA25574;  
AC  
XX 19-JUN-2003 (first entry)  
DT  
XX Prokaryotic essential gene #7231.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
XX Burkholderia fungorum.  
OS

XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Mail D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU21704.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 13444; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 5213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1035 BP; 235 A; 230 C; 306 G; 204 T; 0 U; 0 Other;

Alignment Scores: 242 Length: 1035  
Pred. No.: 8.00 Matches: 8  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 10.13% Gaps: 0  
DB: 7

US-09-972-032-2 (1-79) x AC25574 (1-1035)

QY 48 SerArgTTPProArgSerAlaser 55  
|||||

DB 836 TCGAGATGCCCGAGGTCTGCCAGT 813

RESULT 10  
AAH48630/c  
ID AAH48630 standard; DNA; 1037 BP.  
XX AAH48630;  
XX 21-SEP-2001 (first entry)  
XX Human MLP promoter fragment.  
XX MLP; human; mutation; muscle-specific promoter; cardiovascular disease;  
XX dilative cardiomyopathy; cardiac; gene therapy; myocardial disease;  
XX sarcomer; dystrophin; cardiac actin; hypertrophic cardiomyopathy;  
XX long QT syndrome; chromosome 1p15.1; promoter; ds.  
XX Homo sapiens.  
XX WO200157208-A2.  
XX 09-AUG-2001.  
XX 01-FEB-2001; 2001WO-EP001042.  
XX 03-FEB-2000; 2000DE-01004857.  
XX (SCHD ) SCHERING AG.  
XX Knoell R;  
XX WPI; 2001-483436/52.  
XX New nucleic acid encoding mutant MLP, useful for diagnosis and treatment  
XX of myocardial disease, particularly dilative cardiomyopathy.  
XX Claim 35; Page 51; 53pp; German.  
XX This invention describes a novel nucleic acid (I) encoding an MLP (not  
XX defined) which has a 1273 base pair (bp) sequence (1) that includes a  
XX mutation at base 10 in exon 2 or the third position of codon 112 in exon  
XX 4, is new. The product of the invention has cardiac activity and can be  
XX used for gene therapy. (I), and related nucleic acids or probes, are used  
XX in diagnosis of and/or screening for myocardial diseases (or  
XX predisposition), especially dilative cardiomyopathy. Both specified  
XX mutations are associated with development of these diseases. Antibodies  
XX (Ab) raised against MCP and other peptides encoded by (I) can be used  
XX similarly. Also the regulatory region (III) of the genomic MLP sequence  
XX (optionally when incorporated into vectors or cells) is used in gene  
XX therapy, specifically for prevention and/or treatment of cardiovascular  
XX disease, particularly those which involve a point mutation in a gene  
XX encoding sarcomer, dystrophin or cardiac actin, e.g. hypertrophic  
XX cardiomyopathy, long QT syndrome and dilative cardiomyopathy. The  
XX regulatory region of the MLP gene provides muscle-specific gene  
XX expression. This sequence represents a promoter fragment derived from the  
XX human MLP described in the invention

XX SQ Sequence 1037 BP; 285 A; 232 C; 232 G; 288 T; 0 U; 0 Other;

Alignment Scores: 242 Length: 1037  
Pred. No.: 8.00 Matches: 8  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 10.13% Gaps: 0  
DB: 4

US-09-972-032-2 (1-79) x AAH48630 (1-1037)

QY 26 LeuGlyGluGlyGlnGlyGly 33  
|||||

DB 910 CTGGGGAGAGGCGAGGAGGAGGC 887

```

RESULT 11
ABQ73845
ID ABQ73845 standard; DNA; 1230 BP.
XX
AC ABQ73845;
XX
DT 08-OCT-2002 (first entry)
XX
DE Rhizobium phytase nucleotide sequence SEQ ID NO:7.
XX
KW Rhizobium; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed; gene;
KW ds.
XX
OS Rhizobium sp.
XX
FH Key Location/Qualifiers
CDS 1..1230
FT /*tag= a
FT /EC_number= "3.1.3.8"
FT /product= "phytase"
XX
PN WO200248332-A2.
XX
PD 20-JUN-2002.
XX
PF 12-DEC-2001; 2001WO-US048774.
XX
PR 12-DEC-2000; 2000US-0255090P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
XX
DR WPI; 2002-583504/62.
DR P-PSDB; ABP51932.
XX
PT Novel recombinant phytase protein and polynucleotide for improving
PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.
XX
PS Claim 1; Fig 5G; 208pp; English.
XX
CC The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence encodes a phytase from the present
CC invention
XX
SQ Sequence 1230 BP; 182 A; 434 C; 419 G; 195 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 284 Length: 1230
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 6 Gaps: 0

```

---

```

US-09-972-032-2 (1-79) x ABQ73845 (1-1230)
Qy 47 AlaSerArgTrpProArgSerAla 54
Db 725 GCAAGCCGATGCCCGAGTCGGCT 748
RESULT 12
ACA25997
ID ACA25997 standard; DNA; 1239 BP.
XX
AC ACA25997;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #7654.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Burkholderia mallei.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SRP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU22127.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 13867; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,

```

CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1239 BP; 163 A; 419 C; 480 G; 177 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 286 Length: 1239  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 7 Gaps: 0

US-09-972-032-2 (1-79) x ACA25997 (1-1239)

QY 48 SerArgTrpProArgSerAlaSer 55  
 DB 95 AGCAGATGCCCGCAGCGCTTCA 118

RESULT 13  
 ABQ73844  
 ID ABQ73844 standard; DNA; 1266 BP.

XX ABQ73844;

DT 08-OCT-2002 (first entry)

DE Yersinia pestis phytase nucleotide sequence SEQ ID NO:5.

KW Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed;  
 Gene; ds.

OS Yersinia pestis.

Key Location/Qualifiers  
 CDS 1..1266  
 FT /\*tag= a  
 FT /EC\_number= "3.1.3.8"  
 FT /product= "phytase"

XX WO200248332-A2.

XX 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048774.

XX 12-DEC-2000; 2000US-0255090P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX WPI; 2002-583504/62.

XX P-PSDB; ABP51931.

XX Novel recombinant phytase protein and polynucleotide for improving  
 PT nutritional value of phytate-containing foodstuff, in animal feed and  
 PT feed supplements and to degrade excess phytase from environment or  
 PT sample.

PS Claim 1; Fig 5E; 208pp; English.

XX The present invention describes an isolated phytase protein (I). (I) can  
 CC be used for improving the nutritional value of a phytate-containing  
 CC foodstuff, by contacting the phytate-containing foodstuff with (I), where  
 CC the enzyme catalyses the liberation of inorganic phosphate from the  
 CC phytate-containing foodstuff, and so improving the nutritive value of the  
 CC contacted foodstuff. The liberation of inorganic phosphate occurs prior  
 CC to or after the ingestion of phytate-containing foodstuff by a recipient  
 CC organism. Nucleotide sequences (II) encoding (I) can be used for

CC producing an animal feed, by transforming a plant, plant portion or plant  
 CC cell with a nucleic acid expression vector, comprising (II), culturing  
 CC the plant, plant portion or plant cell under conditions in which the  
 CC phytase protein is expressed and converting the plant, plant portion or  
 CC plant cell into a composition suitable for animal feed. The animal is  
 CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
 CC animal feed and feed supplements as well as in treatments to degrade or  
 CC remove excess phytate from the environment or a sample. (I) reduces  
 CC phytate levels in animal manure and so reduces the phosphate pollution of  
 CC the environment. The present sequence encodes a phytase from the present  
 CC invention

SQ Sequence 1266 BP; 194 A; 444 C; 426 G; 202 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 292 Length: 1266  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x ABQ73844 (1-1266)

QY 47 AlaSerArgTrpProArgSerAla 54

DB 761 GCAGCCGATGCCCGAGGCTCGGCT 784

RESULT 14

AAQ06668

ID AAQ06668 standard; DNA; 1349 BP.

XX AAQ06668;

DT 01-MAR-1991 (first entry)

DE Mycomacterium derived alpha-antigen gene.

KW Vaccine; tuberculosis; HIV-1; ds.

XX Mycobacterium kansasii.

Key Location/Qualifiers  
 CDS 270..1244  
 FT /\*tag= a  
 FT sig\_peptide 270..389  
 FT /\*tag= b

XX EP400973-A.

XX 05-DEC-1990.

XX 30-MAY-1990; 90EP-00305849.

XX 31-MAY-1989; 89JP-00135855.

XX 16-MAR-1990; 90JP-00084310.

XX (AJIN ) AJINOMOTO KK.

XX Matsuo K, Yamaguchi R, Yamazaki A, Yamada T;

XX WPI; 1990-363461/49.

XX P-PSDB; AAR08099.

XX New mycobacterial secretory vector - used to transform host cells, and  
 PT vaccine comprising transformant.

XX Disclosure; Fig 1; 19pp; English.

XX The sequence may be used to produce a recombinant protein comprising  
 CC alpha-antigen and B-cell epitope of the HIV-1 gag antigen p17. The  
 CC product is useful in vaccine production eg. tuberculosis. A live vaccine  
 CC such as M.bovis BCG or M.smegmatis has a low toxicity and long lasting



CC activity  
 SQ Sequence 1349 BP; 232 A; 434 C; 438 G; 245 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 309 Length: 1349  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-972-032-2 (1-79) x AAQ06668 (1-1349)  
 Qy 51 ProArgSerAlaSerArgTrpPro 58  
 Db 751 CCGCGTCCGCATCTCGATGCCG 774

RESULT 15  
 ABQ79295  
 ID ABQ79295 standard; cDNA; 1394 BP.  
 XX  
 AC ABQ79295;  
 XX  
 DT 15-NOV-2002. (first entry)  
 XX  
 DE Mycobacterium-originated alpha antigen encoding sequence.  
 XX  
 KW Antiallergic; Dermatological; Antiasthmatic; Antiinflammatory;  
 KW Ophthalmological; Mycobacterium-originated alpha antigen;  
 KW allergic disease; atopic dermatitis; asthma; allergic rhinitis;  
 KW allergic conjunctivitis; interleukin-4; Th2-type cytokine;  
 KW IGE production; histamine; eosinophilic infiltration; gene; ss.  
 XX  
 OS Mycobacterium kansasii.  
 XX  
 FN WO20026055-A1.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 20-FEB-2002; 2002WO-JP001459.  
 XX  
 PR 20-FEB-2001; 2001JP-00043291.  
 XX  
 PA (PRIM-) PRIMMUNE CORP INC.  
 XX  
 FI Yasutomi Y, Mizutani H;  
 XX  
 DR WP1; 2002-667038/71.  
 XX  
 PT Medicinal use of Mycobacterium-originated alpha-antigen or its gene in  
 PT treating allergic diseases e.g. atopic dermatitis, asthma, allergic  
 PT rhinitis and allergic conjunctivitis.  
 XX  
 PS Disclosure; Page 26-27; 34pp; Japanese.  
 XX  
 CC This invention relates to pharmaceutical compositions for preventing or  
 CC treating allergic diseases containing Mycobacterium-originated alpha  
 CC antigens. The compositions are antiallergic, dermatological,  
 CC antiasthmatic, antiinflammatory and ophthalmological. The antigen and its  
 CC encoded gene are for producing drug compositions in treating allergic  
 CC diseases e.g. atopic dermatitis, asthma, allergic rhinitis and allergic  
 CC conjunctivitis. The drug compositions are made from the protein that can  
 CC inhibit production of interleukin-4, ameliorate Th2-type cytokine-  
 CC predominate immune state and relieve symptoms of allergic diseases such  
 CC as IGE production, histamine release and eosinophilic infiltration. The  
 CC present sequence represents the encoding sequence for a mycobacterium-  
 CC originated alpha antigen. This encoding sequence does not encode the  
 CC protein featured (ABB8362) despite being said to do so in the  
 CC specification  
 XX

SQ Sequence 1394 BP; 241 A; 452 C; 449 G; 252 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 319 Length: 1394  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.13% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-972-032-2 (1-79) x ABQ79295 (1-1394)  
 Qy 51 ProArgSerAlaSerArgTrpPro 58  
 Db 751 CCGCGTCCGCATCTCGATGCCG 774  
 Search completed: July 13, 2004, 15:05:28  
 Job time : 384 secs

This Page Blank (uspio)